

THIS WEEK

EDITORIALS

WORLD VIEW The global climate target of a 2 °C rise is not enough **p.7**

SLOW GOING Introduced tortoises spread ebony seeds again **p.8**



RING CYCLE Chemistry breakthrough could make more molecules **p.9**

Storm warning

Political hostility over global-warming policy in the United States is causing collateral damage. Plans for a National Climate Service deserve better.

Consider it as a shot across the bow. Republicans on the US House Committee on Science, Space, and Technology managed to include language in last month's agreement for fiscal 2011 that stops the National Oceanic and Atmospheric Administration (NOAA) from spending on a new National Climate Service. The temporary restriction has little immediate impact, given that NOAA proposed how to create the service in its 2012 budget request, which is currently up for debate. But the administration of President Barack Obama must now re-engage with lawmakers and make its case for the service, while ensuring that the proposal is not sunk by unrelated partisan battles.

The idea is simple and worthwhile. NOAA wants to collect various climate research and reporting activities under a single umbrella, which it says will make the government machine operate more efficiently and improve the quality of data released to the public — everything from the results of satellite monitoring and climate models to regional forecasts of drought and floods. Months before the spate of storms in April hammered midwestern and southern states, for example, NOAA warned of a higher likelihood of flooding and extreme weather associated with a La Niña circulation in the Pacific Ocean.

House science chairman Ralph Hall (Republican, Texas) has raised concerns about moving forward without a thorough review on Capitol Hill, but a Congress-commissioned external review by the National Academy of Public Administration endorsed the reorganization in September 2010. And Congress will weigh in throughout the budget process. Hall's claims that the creation of a climate service could undermine core research at the agency are plain wrong. NOAA's Office of Oceanic and

Atmospheric Research would see its budget cut by more than half, but that does not mean research is being axed. Nor is NOAA proposing anything new and grandiose at this point. The agency would merely be shifting many of its climate-related activities into a climate service.

Somehow this has become a partisan issue — 227 Republicans voted to approve a similar amendment to bar spending on the climate service during the appropriations debate back in February. It seems that many are determined to conflate the word 'climate' with the contentious debate over global-warming policy.

"Many are determined to conflate the word 'climate' with the debate over global warming."

One of NOAA's core functions is to provide basic — and non-partisan — information on weather and climate, useful for everybody from scientists and governments to farmers, commuters and businesses. Indeed, so valuable is this information that the data themselves have become a commodity to be repackaged and sold on by private companies. The proposed reorganization would improve this service, and appropriators and lawmakers on both sides should endorse it.

Then they should focus on a bigger issue: satellite funding. This year's budget denied the first half of a two-year increase of nearly \$1.2 billion for the Joint Polar Satellite System, threatening a lapse in data and less-accurate forecasting. Building on its long-term prediction, and using satellite data, NOAA accurately forecast April's extreme weather several days in advance. The storms, which still killed hundreds of Americans, are a warning worth heeding. ■

Flagship funding

The European Union plans to throw serious money at serious problems.

The European Commission this week launches six pilots for its multi-billion-euro Future and Emerging Technologies Flagships programme, under the slogan 'science beyond fiction' (see *Nature* doi:10.1038/news.2011.143; 2011).

The programme is, by a considerable margin, the most expensive ever set up in Europe purely for academic consortia. The pilots have been awarded €1.5 million (US\$2.2 million) each for one-year feasibility studies. Two or three will go on to win a colossal €1 billion in funding over ten years.

The science behind the flagship projects really is beyond fiction. The research is designed to address problems that we can foresee but don't yet know how to solve. How will we store the already overwhelming

amounts of data we continue to generate? How can we build better, greener computers and robots? The funded projects will also focus on social or political priorities for the European Union (EU), such as dealing with an ageing society, or monitoring the environmental impact of human activities. Perhaps we will see perceptive robots built to befriend the lonely.

The funding could also be described as beyond fiction; the promised money has yet to be magicked up. The commission clearly hopes that once the projects are fleshed out, they will prove irresistible to the European Parliament and Council of Ministers who must support long-term financing. And the financing is beyond fiction too: the consortia must provide half of the funds themselves, so are relying on being able to mobilize the required half-billion euros from national research agencies, industry or other sources. That's not something that academics have much experience in doing — and, as they will discover, it's not easy to exact long-term commitments for such high-risk research.

The grand EU flagships experiment is itself high risk, but wise. There can be no real losers: all of the consortia plan to continue their work, whether or not their pilots are selected for funding by the commission. Beyond that, who knows? ■



2 °C or not 2 °C? That is the climate question

Targets to limit the global temperature rise won't prevent climate disruption. Tim Lenton says that policy-makers should focus on regional impacts.

As a scientist who works on climate change, I am not comfortable with recommending policy. Colleagues frown on it, and peer review of scientific papers slams anything that could be construed as policy prescription. Yet climate science is under scrutiny in multiple arenas, and climate scientists have been encouraged to engage more openly in societal debate.

I don't want to write policies, but I do want to ensure that global efforts to tackle the climate problem are consistent with the latest science, and that all useful policy avenues remain open. Ongoing negotiations for a new climate treaty aim to establish a target to limit the global temperature rise to 2 °C above the average temperature before the industrial revolution. But that is not enough.

The target is linked to the United Nations Framework Convention on Climate Change (UNFCCC), which aims to "prevent dangerous anthropogenic interference with the climate system". But that noble objective is nearly 20 years old and is framed too narrowly, in terms of the "stabilization of greenhouse gas concentrations in the atmosphere". Long-term goals to limit temperature or concentrations have so far failed to produce effective short-term action, because they do not have the urgency to compel governments to put aside their own short-term interests.

Global average warming is not the only kind of climate change that is dangerous, and long-lived greenhouse gases are not the only cause of dangerous climate change. Target setters need to take into account all the factors that threaten to tip elements of Earth's climate system into a different state, causing events such as irreversible loss of major ice sheets, reorganizations of oceanic or atmospheric circulation patterns and abrupt shifts in critical ecosystems.

Such 'large-scale discontinuities' are arguably the biggest cause for climate concern. And studies show that some could occur before global warming reaches 2 °C, whereas others cannot be meaningfully linked to global temperature.

Disruption of the south- or east-Asian monsoons would constitute dangerous climate change, as would a repeat of historic droughts in the Sahel region of Africa or a widespread dieback of the Amazon rainforest. These phenomena are not directly dependent on global average temperature, but on localized warming that alters temperature gradients between regions. In turn, these gradients are influenced by uneven distribution of anthropogenic aerosols in the atmosphere.

Equally, an abrupt shift in the regions in which dense masses of water form in the North Atlantic could dangerously amplify sea-level rises along the northeastern seaboard of the United States. But the point at which that will occur depends on

the speed of climate change more than its magnitude.

Even when a threshold can be directly related to temperature, as with the melting of ice sheets, it is actually the net energy input that is important. The rapid warming of the Arctic in recent years is attributable less to increasing carbon dioxide levels than to reductions in emissions of sulphate aerosols (which have a cooling effect), and to increases in levels of warming agents, including black-carbon aerosols and the shorter-lived greenhouse gases methane and tropospheric ozone.

Ultimately, crucial climate events are driven by changes in energy fluxes. However, the one metric that unites them, radiative forcing, is missing from most discussions of dangerous climate change. Radiative forcing measures the change in the net imbalance of energy that enters and leaves the lower atmosphere; it is a better guide to danger than

greenhouse-gas concentrations or global warming. It takes into account almost all anthropogenic activities that affect our climate, including emissions of methane, ozone-producing gases and hydrofluorocarbons, and changes in land use and aerosol levels.

I suggest that the UNFCCC be extended. The climate problem, and the political targets presented as a solution, should be aimed at restricting anthropogenic radiative forcing to limit the rate and gradients of climate change, before limiting its eventual magnitude.

How would this help? A given level of radiative forcing is reached long before the resulting global temperature change is fully realized, which brings urgency to the policy process. The 2 °C target would translate into a radiative forcing of about 2.5 Watts per square metre (W m^{-2}), but

to protect major ice sheets, we might need a tougher global target of 1.5 W m^{-2} . We will still need a binding target to limit long-term global warming. And because CO_2 levels remain the most severe threat in the long term, a separate target could tackle cumulative carbon emissions. But while we wait for governments to reach an agreement on CO_2 , we can get to work on shorter-lived radiative-forcing agents.

The beauty of this approach is that it opens separate policy avenues for different radiative-forcing agents, and regional treaties to control those with regional effects. For example, hydrofluorocarbons emissions could be tackled under a modification of the 1987 Montreal Protocol, which aimed to halt ozone depletion. And emissions of black-carbon aerosols and ozone-producing gases could be regulated under national policies to limit air pollution. This would both break the political impasse on CO_2 and help to protect vulnerable elements of the Earth system. ■

Tim Lenton is professor of Earth system science in the College of Life and Environmental Sciences, University of Exeter, UK.
e-mail: t.m.lenton@exeter.ac.uk

GLOBAL
AVERAGE
WARMING
IS NOT THE ONLY KIND
OF CLIMATE CHANGE
THAT IS
DANGEROUS.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/1yurph

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

NANOTECHNOLOGY

Bigger screens with nanotubes

Visual displays based on organic light-emitting diodes (OLED) promise to be lighter and brighter than those using older liquid crystal display (LCD) technology. But the polycrystalline silicon transistors that power OLED screens cannot be made in uniform size and shape, causing variation from one pixel to another and limiting display size.

Andrew Rinzler at the University of Florida in Gainesville and his group have created a transistor in which the 'source' electrode is made from a single-layer carbon nanotube network. These transistors can be incorporated into devices made with a wide range of organic materials to provide the required currents, potentially permitting the manufacture of larger screens operating at a voltage comparable to that of silicon-based OLEDs. The resulting devices consume eight times less power than those based on previous technologies and can, theoretically, prolong OLED lifetime by a factor of four.

Science 332, 570–573 (2011)

ECOLOGY

Understudy takes on tortoise's role

A controversial approach to ecosystem conservation — replacing extinct species with functionally similar ones from elsewhere — has been successfully demonstrated on a tiny island in the Indian Ocean.

The ebony tree (*Diospyros egrettarum*) was unable to rebound after extensive logging on Ile aux Aigrettes because the giant tortoises and skinks that used to eat its fruit and

disperse its seeds had become extinct. So Christine Griffiths at the University of Bristol, UK, and her colleagues introduced 19 adults of the Aldabra giant tortoise (*Aldabrachelys gigantea*, pictured) from another island between 2000 and 2009. The animals promptly began dispersing ebony seeds. Seeds that had passed through the digestive



tracts of tortoises germinated more often and faster than those that had not. Ebony seedlings now dot the island. **Curr. Biol.** doi:10.1016/j.cub.2011.03.042 (2011)

GENOMICS

A guided tour of the genome

Help is at hand for scientists struggling to make sense of the current flood of human genome sequence data: the Encyclopedia of DNA Elements (ENCODE). Now an accompanying user guide is available.

The guide — published by a consortium composed

of dozens of international research groups — describes data from more than 100 human cell types that define the functional elements in the genome, including more than 2 million regulatory regions. The team mapped RNA transcribed from DNA; protein-binding sites; and 'epigenetic' modifications to DNA's structure, such as DNA methylation patterns. Together, these should help researchers work out possible roles for sequence variants that have been linked to a disease.

For example, ENCODE data helped to clarify how a DNA region upstream of a cancer-promoting gene called *c-Myc* regulates the gene:

Jellyfish eyes on the sky

Box jellyfish seem to have eyes that peer constantly upwards, allowing them to navigate by detecting features on land.

Anders Garm at the University of Copenhagen and his colleagues made video recordings of the box jellyfish *Tripedalia cystophora* (pictured), which has a total of 24 eyes, made up of four types. The team found that four of these — the 'upper lens' eyes — always point straight upwards, regardless of the animals' orientation.

T. cystophora that were moved away from their preferred habitat near mangroves rapidly swam back, unless they were moved farther than 8 metres away. At this distance, surface ripples and the eyes' limited resolution would cripple the jellyfish's view of the mangrove canopy. Blocking the canopy from view with a white sheet also left the jellyfish swimming in random directions.

Curr. Biol. doi:10.1016/j.cub.2011.03.054 (2011)



ZOOLOGY

J. BIELECKI

by attracting and binding proteins that enhance its expression.

PLoS Biol. 9, e1001046 (2011)

ORGANIC CHEMISTRY

Fusing rings with fluorine

The Friedel–Crafts reaction is one of the most beloved and well used in the organic chemist's recipe book, allowing the attachment of chemical groups to aromatic rings through a carbon–carbon bond. Jay Siegel and his colleagues at the University of Zurich in Switzerland have now devised a way to attach one aromatic ring to another, which they say will allow the synthesis of more complex compounds.

Using a silicon-based reagent and an acid catalyst, the authors activated the normally stable bond between a fluorine atom and one of the carbon atoms of an aromatic ring. This led, in turn, to the formation of a new bond between the carbon and another carbon atom on a different aromatic ring. Meanwhile, the acid catalyst is regenerated to drive another cycle of the reaction.

Science 332, 574–577 (2011)

NEUROSCIENCE

Monkey recalls what monkey saw

Humans can remember information from the past, such as the appearance of a childhood home, but attempts to test this ability in other animals have been hampered by a lack of good testing methods. Now Benjamin Basile and Robert Hampton of Emory University in Atlanta, Georgia, have designed a touchscreen computer task for rhesus macaques and used it to show that the primates can recall simple shapes from memory.

Five male monkeys trained on the computer task were able to fill in blanks on a grid to reproduce previously viewed two- and three-square shapes, demonstrating recall. The finding may provide a new

animal model for memory studies and suggests that a common ancestor with humans came under selection pressure for this detailed and flexible use of memory. *Curr. Biol.* doi:10.1016/j.cub.2011.03.044 (2011)

IMMUNOLOGY

Worm-proofing the gut

Pathogenic worms have more than the immune system to contend with in mammals, with a mucus-forming protein also mounting a defence. Sugar-coated mucin proteins form a thick protective mucus layer over organs such as the gut and lungs. David Thornton and Richard Grencis at the University of Manchester, UK, and their team report that the mucin MUC5AC also directly lowers the viability of a gut-dwelling nematode worm.

MUC5AC normally occurs in the lungs, but intestinal levels shoot up in mice infected with *Trichuris muris* worms, a close relative of a nematode that afflicts humans. Mice lacking the *Muc5ac* gene had chronic worm infections, despite showing strong immune responses. In *T. muris* cultured with human cells, the protein limited the worms' production of the cellular energy molecule ATP — a sign of viability. *J. Exp. Med.* doi:10.1084/jem.20102057 (2011)

OCEANOGRAPHY

A tsunami's trip around the world

Rather than travelling straight across the Pacific Ocean, the Sumatran tsunami of December 2004 (pictured) took a roundabout route. It headed southwards towards Antarctica before looping back up to arrive at the northwest coast of North America. The 22,000-kilometre-long trip followed mid-oceanic ridges.

Alexander Rabinovich of the Canadian Department of Fisheries and Oceans in Sidney, British Columbia, and his

COMMUNITY CHOICE

The most viewed papers in science

EVOLUTIONARY GENETICS

Hunting for birth-timing genes

HIGHLY READ
on plos.org
in April

A gene has been linked to preterm birth using an approach that focuses on genes that have evolved faster in humans than in other primates.

Louis Muglia at Vanderbilt University in Nashville, Tennessee, and his colleagues compared gestation times relative to neonatal size across 20 primate species and found that humans have the shortest gestation time — probably to make the delivery of large-headed babies through narrow birth canals easier. For this to happen, the researchers reasoned that genes regulating birth timing have probably evolved faster in humans. So they looked for genes showing signs of accelerated evolution along the human lineage.

The authors selected 150 genes and analysed their surrounding regions in the genomes of 328 mothers, teasing out one gene, *FSHR*, with the strongest link to preterm birth. *PLoS Genet.* 7, e1001365 (2011)

co-workers analysed data from pressure sensors deep in the northeast Pacific Ocean. The team detected tsunami waves at the first sensors around 34–35 hours after the earthquake and inferred the waves' direction by analysing their time of arrival at different sensors.

The waves continued for 3.5 days, suggesting that tsunamis produce a long-lasting energy flux that is conserved and transmitted by oceanic ridges. *Geophys. Res. Lett.* doi:10.1029/2011GL047026 (2011)



METABOLISM

Diabetes drug affects the brain

The diabetes drug rosiglitazone improves insulin sensitivity by acting not only on fat and muscle cells, but also on the brain.

Rosiglitazone binds to a protein receptor called PPAR- γ , which regulates lipid and glucose metabolism. One side effect, however, is weight gain. Randy Seeley and his colleagues at the University of Cincinnati in Ohio found that the drug activates PPAR- γ in the rat central nervous system. Administering rosiglitazone directly to the brain, or overexpressing PPAR- γ in a brain region called the hypothalamus, boosted the rats' appetite and weight gain.

Meanwhile, Jerrold Olefsky at the University of California, San Diego, Michael Schwartz at the University of Washington in Seattle and their colleagues found that mice lacking PPAR- γ in neurons ate less food and used more energy than normal mice. But rosiglitazone didn't work in these mice — it seems that active neuronal PPAR- γ is required for the drug's insulin-sensitizing effects.

Nature Med. doi:10.1038/nm.2349; doi:10.1038/nm.2332 (2011)

NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch

AFP/GETTY

SEVEN DAYS

The news in brief

POLICY

Canada's election

National elections in Canada on 2 May brought bad news for environmentalists, even though the Green Party won its first-ever parliamentary seat. After five years of minority rule, the Conservative Party won an outright majority; the party is generally hostile to efforts to address climate change, and is enthusiastic about extracting oil from western Canada's tar sands. See go.nature.com/wlyicg for more.

Scientist glut?

Amid concerns over the increasing demand for grants and the length of time it takes to train a scientist, the US National Institutes of Health (NIH) has asked a panel of external advisers to report on what a future biomedical-research workforce should look like. It will tackle questions such as how many scientists the United States needs, and how the country should train them. The group, named on 27 April, will make its recommendations to NIH director Francis Collins' advisory committee, possibly by summer 2012. See go.nature.com/rfowsx for more.

Biodiversity plan

On 3 May, the European Commission published a new plan to improve Europe's biodiversity over the next decade. The strategy includes targets on sustainable agriculture, safeguarding fish stocks, controlling invasive species and protecting and restoring ecosystems. It falls roughly into line with agreements made at a biodiversity summit in Nagoya, Japan, last October (see *Nature* **468**, 14; 2010) and targets agreed in March

2010 by the Council of the European Union to halt biodiversity loss by 2020.

BUSINESS

Hepatitis advance

Two new drugs against the hepatitis C virus (HCV) have won unanimous votes of confidence from advisers to the US Food and Drug Administration (FDA). On 27 April, the FDA's Antiviral Drugs Advisory Committee recommended approving boceprevir, developed by Merck, headquartered in Whitehouse Station, New Jersey. The next day, the same committee gave its support to telaprevir, developed by Vertex Pharmaceuticals in Cambridge, Massachusetts.

Both drugs block HCV's protease enzyme. If approved by the FDA, they would be the first therapeutics on the market to directly target HCV, which is currently treated with general antivirals and immune-boosting proteins. See go.nature.com/qjoqfn for more.

Health-care buyout

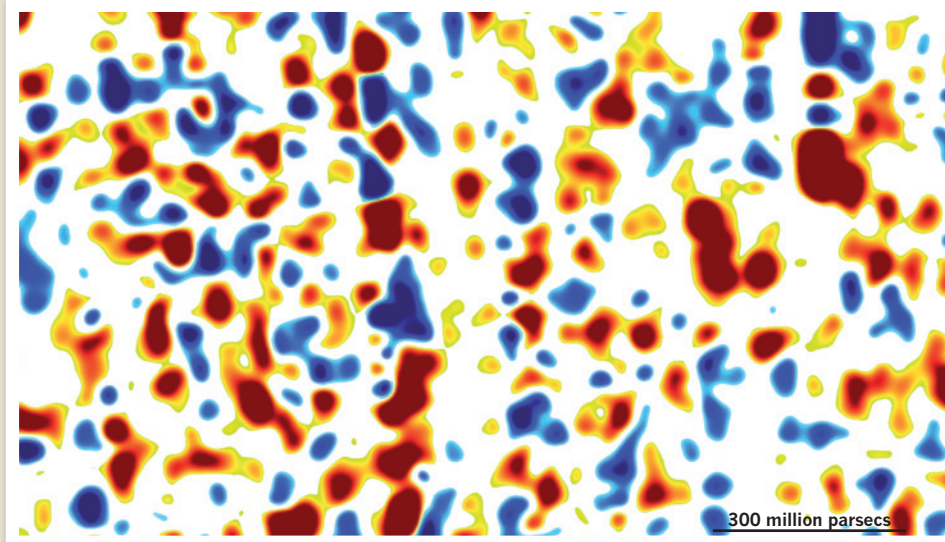
US health-care giant Johnson & Johnson will take over medical-device manufacturer Synthes in a deal worth 19 billion Swiss francs (US\$21.8 billion), the companies announced on 27 April. Synthes, headquartered in Solothurn, Switzerland, makes implants, biomaterials and instruments for orthopaedic surgery.

Pharma takeover

Teva Pharmaceuticals, the world's largest generic drug maker, headquartered near Tel Aviv, Israel, announced on 2 May a US\$6.8-billion deal to buy Cephalon, which makes the narcolepsy treatment modafinil. Cephalon, based in Frazer, Pennsylvania, had been fighting off a hostile \$5.7-billion takeover bid from Canadian firm Valeant Pharmaceuticals of Mississauga, Ontario.

Stem-cell joy

Shares in several stem-cell firms rallied in the wake of a US appeals court decision to overturn an injunction that would freeze federal funding for research on human embryonic stem cells. After



A. SLOSAR/BOSS LYMAN-ALPHA COSMOLOGY WORKING GROUP

Mapping the distant Universe

The first three-dimensional map of the distant Universe, showing clumps of hydrogen gas between 3 billion and 3.7 billion parsecs away, was released on 1 May at a meeting of the American Physical Society in Anaheim, California. The map — the fruits of the Baryonic Oscillation Spectroscopic Survey experiment — was made by measuring

14,000 quasars, the luminous nuclei of early galaxies. Their light is absorbed at particular wavelengths as it passes through the hydrogen. Ripples in this gas (a two-dimensional slice is pictured, with density of gas increasing from blue to red) could shed light on how dark energy drove the expansion of the early Universe. See go.nature.com/fromkf for more.

UNIV. MICHIGAN

the 29 April ruling, StemCells, Advanced Cell Technology, Pluristem Therapeutics, Aastrom Biosciences and Geron all saw gains. See page 15 for more on the decision.

PEOPLE

Russian space chief

Russia's prime minister, Vladimir Putin, has fired the chief of the nation's space agency, Anatoly Perminov. Perminov, who has headed Roscosmos since March 2004, will be replaced by deputy defence minister Vladimir Popovkin. A change had been widely rumoured after December 2010, when three satellites for Russia's global navigation network, GLONASS, crashed into the Pacific Ocean on launch.

Lab saboteur

The US Office of Research Integrity last week issued a finding of research misconduct against former University of Michigan postdoctoral fellow Vipul Bhargava. Last year Bhargava was caught on video (**pictured**) sabotaging the work of a student in his lab (see *Nature* **467**, 516–518; 2010). At the time he was ordered to pay more than US\$30,000 by a Michigan court for destroying property. The federal government now says his acts constituted research misconduct because



they resulted in falsified data. Bhargava, now in India, has been barred from receiving US federal funding for three years. See go.nature.com/lexd9a for more.

Research head quits

Roger Beachy has resigned as director of the US National Institute of Food and Agriculture (NIFA) to spend more time with his family. Beachy, an eminent plant biotechnologist who retains a position at Washington University in St Louis, has led NIFA since its inception in 2009. See page 19 for more.

EVENTS

Tornado damage

Powerful storms and tornadoes that devastated the southern United States last week, killing around 350 people, also knocked out power to the Browns Ferry nuclear plant near Athens, Alabama. But the plant's three reactor units, which have a combined capacity of around 3.3 gigawatts, shut down safely

after diesel generators kicked in. See go.nature.com/bwbadj for more.

Fukushima safety

A scientific adviser to Japan's government resigned from his post on 29 April, complaining that the safety limit set for radiation in schools around the Fukushima nuclear plant was an ad hoc measure and not in line with international standards. Toshiso Kosako, a radiation-safety expert at the University of Tokyo, said the government's safety limit — 20 millisieverts a year — was too high. Local parents and lobby groups have demanded that the government set stricter limits.

RESEARCH

Greenland ice

Ice sheets in Greenland may be more stable than previously thought, results from the North Greenland Eemian Ice Drilling (NEEM) project suggest. In the Eemian interglacial period (130,000–115,000 years ago), temperatures were as much as 5°C warmer than today and sea levels rose by up to 7–8 metres. But new ice cores suggest that melting of the Greenland ice sheet at that time caused global sea levels to rise by only 1 or 2 metres; the remainder may have been made up by ice loss from Antarctica. The findings were presented on 28 April at a symposium at the University

COMING UP

9–13 MAY

The European Materials Research Society teams up with its US counterpart to hold the first bilateral conference on energy, in Nice, France.

go.nature.com/tgxfsi

12 MAY

Ministers from eight Arctic nations discuss how to manage the region at the Arctic Council's biennial meeting in Greenland.

go.nature.com/zeeks2

10–13 MAY

The workings and governance of the Intergovernmental Panel on Climate Change are up for review at the panel's 33rd general assembly, in Abu Dhabi.

go.nature.com/utykml

of Wisconsin in Madison. See go.nature.com/nqewa6 for more.

EU biology links

Three ambitious biological sciences infrastructure projects, costing €700 million (US\$1 billion), were given the go-ahead in Europe on 3 May. One network will update and link facilities in 26 European countries that maintain collections of key research microbes. Another project will link facilities in ecosystem science, and a third will connect facilities and create data repositories for researchers in systems biology. The new projects, starting in 2014–15, are part of an updated wish list of science facilities drawn up by Europe's leading researchers, the European Strategy Forum on Research Infrastructures. See go.nature.com/dtuwn for more.

➔ **NATURE.COM**

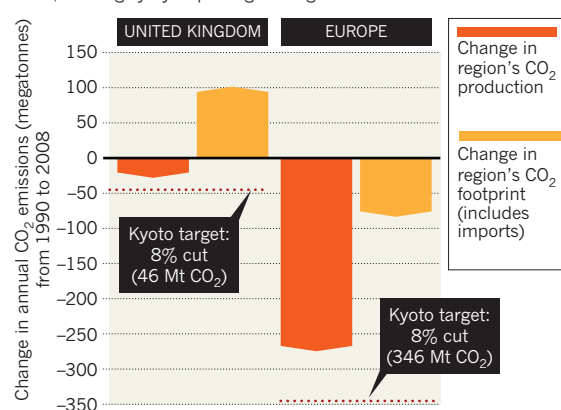
For daily news updates see:
www.nature.com/news

TREND WATCH

Developed nations are responsible for more carbon dioxide emissions than they produce, because they import goods made in other countries. A study of emissions from 113 countries for 1990 to 2008 (G. P. Peters *et al.* *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.1006388108; 2011) shows that developed countries (as classed under the Kyoto Protocol) increased their CO₂ footprint by 7% — even though they reported 2% production cuts. The chart shows the effect for the United Kingdom and Europe.

HOW TRADE AFFECTS CARBON FOOTPRINTS

Rich regions have achieved cuts in carbon emissions since 1990, but largely by importing more goods from elsewhere.



SOURCE: G. PETERS/CICERO

NEWS IN FOCUS

CHINA Ambitious ten-year plan for a space station revealed **p.14**

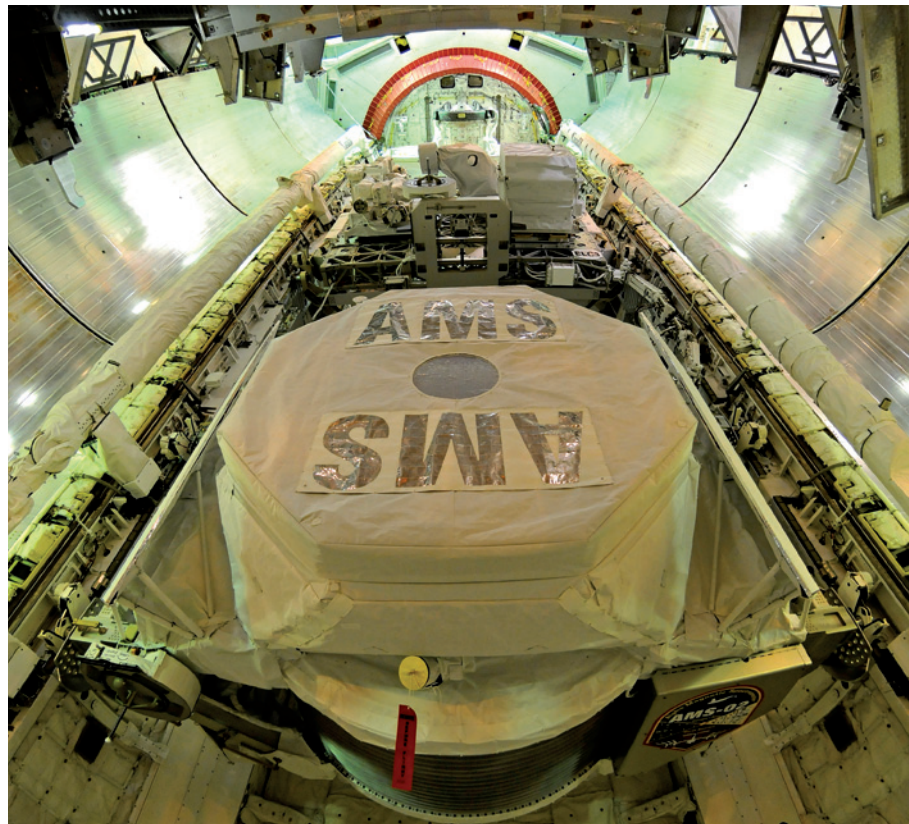
STEM-CELL FUNDING US court decision could dispel the threat of a ban **p.15**

AGRICULTURE A research revolution loses its leader **p.19**

SPACE Is suborbital flight a joy ride or a boon to science? **p.21**



M. FAMILIETTI/AMS COLLABORATION



The Alpha Magnetic Spectrometer will ride to orbit in the space shuttle *Endeavour*'s cargo bay.

ASTROPHYSICS

Antiuniverse here we come

A controversial cosmic-ray detector destined for the International Space Station will soon get to prove its worth.

BY EUGENIE SAMUEL REICH

The next space-shuttle launch will inaugurate a quest for a realm of the Universe that few believe exists.

Nothing in the laws of physics rules out the possibility that vast regions of the cosmos consist mainly of antimatter, with antigalaxies,

antistars, even antiplanets populated with antilife. "If there's matter, there must be antimatter. The question is, where's the Universe made of antimatter?" says Samuel Ting, a Nobel-prizewinning physicist at the Massachusetts Institute of Technology in Cambridge, Massachusetts. But most physicists reason that if such antimatter regions existed,

we would have seen the light emitted when the particles annihilated each other along the boundaries between the antimatter and the matter realms.

No wonder, then, that Ting's brainchild, a US\$2-billion space mission sold partly on the promise of looking for particles emanating from antigalaxies, is fraught with controversy. But the project has other, more mainstream scientific goals. So most critics held their tongues last week as the space shuttle *Endeavour* prepared to deliver the Alpha Magnetic Spectrometer (AMS) to the International Space Station, in a flight delayed by shuttle problems until later this month.

PUSHING THE BOUNDARIES

Seventeen years in the making, the AMS is the product of former NASA administrator Dan Goldin's quest to find remarkable science projects for the space station and of Ting's fascination with antimatter. Funded by NASA, the US Department of Energy and a consortium of partners from 16 countries, it has prevailed despite delays and technical problems, and the doubts of many high-energy and particle physicists.

"Physics is not about doubt," says Roberto Battiston, deputy spokesman for the AMS and a physicist at the University of Perugia, Italy. "It is about precision measurement." As their experiment headed to the launch pad, he and other scientists were keen to emphasize the AMS's unprecedented sensitivity to the gamut of cosmic rays that rain down on Earth. That should allow it not just to detect errant chunks of antimatter from the far Universe, but also to measure the properties of cosmic rays, the high-energy, charged particles flung from sources ranging from the Sun to distant supernovae and γ -ray bursts.

On Earth, cosmic rays can only be detected indirectly, from the showers of secondary particles they produce when they slam into molecules of air high above the ground. From space, the AMS will get an undistorted view. "We'll be able to measure cosmic-ray fluxes very precisely," says collaboration member physicist Fernando Barão of the Laboratory of Instrumentation and

NATURE.COM
For more on Samuel Ting and the AMS, see:
go.nature.com/m9unz8

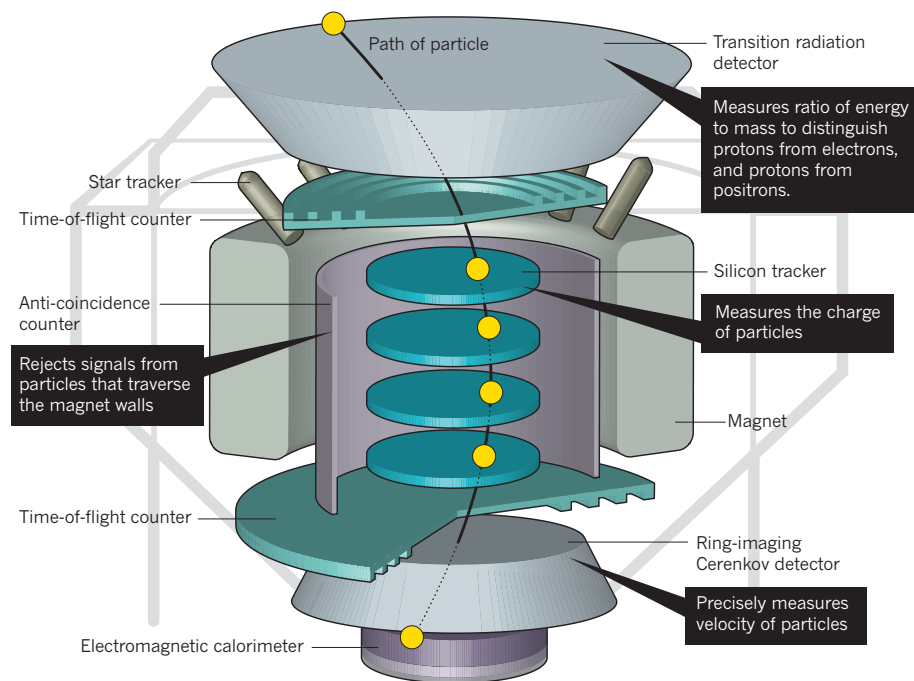
Experimental Particle Physics in Lisbon. "The best place to be is space because you don't have Earth's atmosphere ►

► that is going to destroy those cosmic rays.” No matter what happens with the more speculative search for antimatter, the AMS should produce a definitive map of the cosmic-ray sky, helping to build a kind of astronomy that doesn't depend on light.

The AMS consists of a powerful permanent magnet surrounded by a suite of particle detectors. Over the ten or so years that the experiment will run, the magnet will bend the paths of cosmic rays by an amount that reveals their energy and charge, and therefore their identity.

LOOKING FOR COSMIC CURVEBALLS

The toroidal magnet at the heart of the Alpha Magnetic Spectrometer bends the path of charged, high-speed particles, helping researchers to identify them.



Some will turn out to be heavy atomic nuclei, and any made from antimatter will reveal themselves by bending in the opposite direction from their matter counterparts (see 'Looking for cosmic curveballs').

By counting positrons — antimatter electrons — the AMS could also chase a tentative signal of dark matter, the so-far-undetected stuff that is thought to account for much of the mass of the Universe. In 2009, researchers with the Russian-Italian Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics, flying on a Russian satellite, published evidence of an excess of positrons in the space environment surrounding Earth (O. Adriani *et al. Nature* **458**, 607–609; 2009). One potential source is the annihilation of dark-matter particles in the halo that envelops the Galaxy.

Another speculative quest is to follow up on hints of 'strange' matter, a hypothetical substance, perhaps found in some collapsed stars, that contains strange quarks along with the up and down quarks in ordinary nuclei. NASA's AMS programme manager, Mark Sistilli, says that hints of strange matter were seen in a pilot flight of the AMS aboard the shuttle in 1998, but that the results were too tentative to publish.

Thanks to its status as an exploration mission, the AMS did not need to go through the peer review that NASA would normally require of a science mission. But Sistilli emphasizes that it earned flying colours from committees convened by the energy department, which is supplying \$50 million of the funding. Now their confidence will be put to the test. ■

SPACE FLIGHT

China unveils its space station

Plans for modest outpost solidify 'go it alone' approach.

BY DAVID CYRANOSKI

The International Space Station (ISS) is just one space-shuttle flight away from completion, but the construction boom in low-Earth orbit looks set to continue for at least another decade. Last week, China offered the most revealing glimpse yet of its plans to deploy its own station by 2020. The project seems to be overcoming delays and internal resistance and is emerging as a key part of the nation's fledgling human space-flight programme. At a press briefing in Beijing, officials with the China Manned Space Engineering Office even announced a contest to name the station, a public-relations gesture more characteristic of space programmes in the United States, Europe and Japan.

China first said it would build a space

station in 1992. But the need for a manned outpost "has been continually contested by Chinese space professionals who, like their counterparts in the United States, question the scientific utility and expense of human space flight", says Gregory Kulacki, China project manager at the Union of Concerned Scientists, headquartered in Cambridge, Massachusetts. "That battle is effectively over now, however, and the funds for the space station seem to have been allocated, which is why more concrete details are finally beginning to emerge."

Significantly smaller in mass than the ISS and Russia's Mir space station (see 'Rooms with a view'), which was deorbited in 2001, the station will consist of an 18.1-metre-long core module and two 14.4-metre experimental modules, plus a manned spaceship and a cargo craft. The three-person station will host

scientific experiments, but Kulacki says it also shares the broader goals of China's human space programme, including boosting national pride and China's international standing.

The space-station project will unfold in a series of planned launches over the next ten years. Last Friday, official state media confirmed that the Tiangong 1 and Shenzhou 8 unmanned space modules will attempt a docking in orbit later this year, a manoeuvre that will be crucial for assembling a station in orbit. If that goes well, two manned Shenzhou craft will dock with Tiangong 1 in 2012. China will then move on to proving its space laboratory capabilities, launching Tiangong 2 and Tiangong 3, which are designed for 20-day and 40-day missions, respectively, over the next 3 years. Finally, it will launch the modules that make up the station.



An unusual pilgrimage: a grey whale spotted off the coast of Israel.

MARINE BIOLOGY

Wayward whale not a fluke

Warming Arctic cited as likely cause of freak migration.

BY NADIA DRAKE

The sighting of a lone grey whale (*Eschrichtius robustus*) last year off the beaches of Israel, and then again near Spain, came as a surprise to many. How did a creature normally found in Pacific waters come to be in the Mediterranean Sea? Although no one knows what happened to the bus-sized mammal after its last appearance in May 2010, a group of researchers now suggests that the sighting might indicate a wider trend: the mixing of northern Atlantic and Pacific marine ecosystems, made possible by the climate-driven depletion of Arctic sea ice.

Marine biologist Aviad Scheinin, from the Israel Marine Mammal Research and

Assistance Center in Haifa, and his colleagues considered the errant whale's most likely origin and route. In a paper published online on 19 April in *Marine Biodiversity Records* (A. P. Scheinin *et al.* *Mar. Biodiv. Rec.* **4**, e28; 2011), they rule out a source in the presumed-extinct North Atlantic population. Comparing photos of the whale's fluke with those of individuals in the small, critically endangered western (North) Pacific population, they found no matches, implying that the whale is a member of the roughly 20,000-strong eastern North Pacific population.

After feeding in the Chukchi and Bering seas during the summer months, grey whales normally head south through the Pacific. This one could have followed an Arctic route instead,

perhaps along the Siberian coast where sea ice has been in marked retreat.

"The whale was supposed to go to California or Mexico," Scheinin says. "But it got lost and ended up in the North Atlantic. Then it started to go south, keeping the land on its left as it would if it were travelling down the North American coastline, and made a left at Gibraltar."

In autumn 2009, when the whale presumably would have started its odyssey, sea-ice coverage in the Arctic was sparse enough to make such a passage plausible, says Harry Stern, a mathematician at the University of Washington in Seattle, who studies sea ice. "The opening of the passages that we've seen in the last four or five years is unprecedented," he adds.

John Calambokidis, a research biologist with the Cascadia Research Collective, a non-profit scientific and educational organization in Olympia, Washington, says the authors have done a good job in considering factors such as grey whale populations, feeding habits and swimming speeds. "A grey whale in the Mediterranean does not make sense," he says. "But among the explanations for the bizarre occurrence, this is definitely the most plausible."

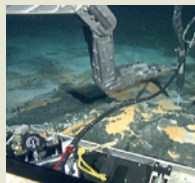
The lack of a tissue sample means that the whale can't be traced to its original population using genetic markers. With no further data, it is premature to conclude that the sighting is related to climate change, says ecologist Kristin Laidre of the University of Washington in Seattle. But climate is sure to affect future whale sightings, she says. "There's no doubt that ice loss will allow the Arctic to act as a corridor for marine species exchange between areas that were previously geographically isolated," she says. "Whales will migrate to the Arctic earlier, move farther north and stay longer. Those are things we predict, and expect to see."

Grey whales aren't the only creatures whose ranges might expand as summer sea ice contracts. "You could make an argument for any species with an open-ocean phase in its life history," says evolutionary biologist David Tallmon, from the University of Alaska Southeast, in Juneau. Potential travellers range from the smallest diatoms to the largest whales — and include terrestrial species seeking colder temperatures nearer the poles (see *Nature* **468**, 891; 2010). "Whole thermal regimes changing could lead to all sorts of weird ecological effects," Tallmon says. ■



**MORE
ONLINE**

TOP STORY



The threat and promise of undersea methane deposits
go.nature.com/q3ca4q

MORE NEWS

- Robotic probe could explore icy region on Earth and beyond go.nature.com/kbn92m
- Overfishing hits all creatures great and small go.nature.com/kazigt
- China's carbon emissions will peak — eventually go.nature.com/zvfooc

ON THE BLOG



How DNA may have confirmed Osama bin Laden's death go.nature.com/mgyqea

POLICY

A how-to for peer review

A guide surveys the range of practices in Europe — and offers suggestions for improvement.

BY ALISON ABBOTT

Your cross-disciplinary grant application has been rejected. Do you wonder if specialist reviewers really grasped its scope? Would you feel more, or less, confident in the process if reviewers were paid?

A report from the European Science Foundation (ESF) offers guidance on the fairest ways to evaluate project proposals. *The European Peer Review Guide*, out last week, maps grant-reviewing practices among European funding agencies and sets out recommendations.

Europe is a kind of laboratory for peer review, with programmes that span disciplines, and countries with varying evaluation criteria, review procedures and incentives for reviewers. There is no single correct model, says co-author Marc Heppener, the ESF's director of science and strategy development. "The report refrains from using the term 'best practice,' and refers instead to 'good practice,'" he says.

The report lists commonly used options for

everything from managing conflicts and confidentiality to organizing documentation. It lays out different methods of selecting experts and provides clear definitions of common terms such as 'excellence' and 'transparency'. And it offers specific advice, such as the need for gender balance among reviewers and for incorporating right-of-reply in the decision-making process. As for paying reviewers, it recommends not doing so unless "really necessary".

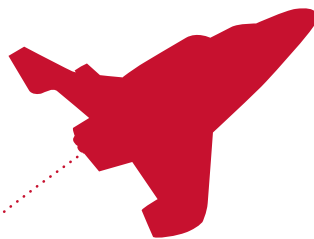
Heppener hopes that the guide will be useful to research agencies around the world. But it was commissioned in 2009 for a narrower purpose by the heads of the European research councils. They realized that better-harmonized peer-review systems would make it easier to share grant-proposal reviews, a pressing need as the number of international research programmes grows. "National research agencies were aware that reviews of a proposal made in one country may not be valid in a second country, and this was frustrating," says Heppener.

As well as giving a how-to guide for peer

review, the report pays special attention to research projects at the intersections of traditional disciplines. It describes ways to assess the quality of all parts of any 'pluridisciplinary' project, and the value of integrating them, with a minimum of peer-review stages.

The European Peer Review Guide is a rolling report, which will be regularly updated as agencies modify their procedures, adopting options that suit their own cultural constraints. Some agencies in large countries, for example, rely mostly on national reviewers, whereas others in small countries may be obliged by law to include international reviewers.

"It's a useful and important report, particularly for those of us who want to fund global projects," says David Stonner, deputy director of the office of international science and engineering at the US National Science Foundation (NSF), who was a formal observer of the procedures for creating the guide. "We need a common language and this report will get a lot of attention at the highest levels of the NSF." ■



ALONG FOR THE RIDE

Reusable commercial rockets will soon be able to take scientists — and tourists — on suborbital spaceflights. Are these vehicles vital research tools, or an expensive dead end?

BY LEE BILLINGS

A

s NASA's space shuttle *Discovery* roared into the sky on 24 February 2011, the bass rumble of its main engines and the staccato crackle of its solid-rocket boosters rolled out across the central Florida countryside, growing fainter and fainter with distance. Viewed from a hotel patio some 65 kilometres away in Orlando, the pillar of flame seemed to rise soundlessly, a silent apparition from a bygone era. *Discovery* was on its final mission; only two shuttle flights were left before the programme ended for good. In the United States, the classical era of the nation's human spaceflight was drawing to a close, 50 years after it began with the 15-minute flight of astronaut Alan Shepard on 5 May 1961.

Three days after *Discovery*'s launch, in the bar of the Orlando hotel, two planetary scientists are talking with a group of fellow researchers about what should come next. Sipping his drink, Daniel Durda laments that after half a century, only about 500 people have flown in space. Access to humanity's final frontier is still restricted to people employed by a handful of powerful governments and corporations, plus the occasional joyriding mega-millionaire. "I'd prefer for anyone to be able to go, for any reason they choose," says Durda, of the Boulder, Colorado, branch of the Southwest Research Institute (SwRI).

His companion, Catherine Olkin, also of the SwRI, agrees. "What we're doing is the next step," she says. "There are huge opportunities up there, not just for science, but for everyone."

That next step is the subject of the meeting that has brought them all here. The second annual Next-Generation Suborbital Researchers Conference (NSRC) is inspired by the growth in recent years of a plethora of commercial companies making rockets designed to carry instruments and paying passengers more than 100 kilometres above Earth — past the edge of the atmosphere and into space. 'Suborbital' denotes vehicles that will come down again without entering orbit, but will still offer researchers precious minutes to make astronomical observations unblurred by the atmosphere, or to study physical processes in the absence of gravity.

Indeed, conference attendees are already buzzing with the news that the SwRI has budgeted US\$1.3 million for a four-year suborbital science

programme, a portion of which will be used to book passenger seats on spacecraft for Durda, Olkin and their colleague Alan Stern, the SwRI's associate vice-president of research and development. If all goes as planned, the three researchers will be flying into space as fully fledged astronauts by mid-2013.

The SwRI is so far the only research institution to have made such a deal, and everyone here knows the arguments for caution. None of the leading suborbital companies has yet flown its vehicles into space and back. And seats on future flights are going for some \$100,000–200,000, yet will give researchers no more than five minutes of weightless 'hang time' above the atmosphere. For most data-collecting needs, it is just as effective to launch automated equipment on an unmanned rocket. Many space scientists, therefore, remain dubious about the usefulness of commercial suborbital spaceflights, particularly those on which researchers accompany their equipment.

But few of those sceptics have made the trip to the Orlando conference, where the prevailing mood is enthusiasm. No one embodies that feeling better than Stern. He hasn't made it to the hotel bar this evening, but that is only because he is busy conferring with launch-industry executives — not to mention preparing to chair panel sessions, deliver a plenary talk and give press conferences.

As the principal investigator for NASA's New Horizons mission to Pluto and former head of the agency's Science Mission Directorate, Stern has a reputation for making big things happen — including the NSRC itself, which he and Durda helped to organize. He frequently compares the state of the nascent suborbital industry to the early days of commercial aviation and personal computing. Eventually, he explains when *Nature* catches up with him the next day, "we'll no longer have one centrally planned space programme where only NASA has the keys to the space shuttle, and everyone has to funnel through that system. Everyone who can afford a ticket will go, and that will generate a lot of innovation, a lot of variety. This is going to be like the Wild West."

Ironically, this 'centrally controlled' programme

"EVERYONE WHO CAN AFFORD A TICKET WILL GO. THIS IS GOING TO BE LIKE THE WILD WEST."

➤ NATURE.COM

For more on NASA's push for commercial spaceflight, visit: go.nature.com/zwdwf2w

SPACE FOR EVERYONE

A multitude of private companies have joined the competition to provide launch services.



1 FALCON 9

Company: SpaceX
Max altitude: Low Earth orbit
Microgravity: Indefinite
Passengers/Pilots: Seven people in capsule
Status: 2010: Rocket and Dragon cargo capsule achieve orbit. 2011: NASA funds Dragon upgrade for humans.

2 SPACESHIP TWO

Company: Virgin Galactic
Max altitude: 110 km
Microgravity: Four minutes
Passengers/Pilots: Six/two
Status: 2010–11: eight tests, four as a free-flying glider. 2012: commercial passenger flights expected.



3 LYNX

Company: XCOR Aerospace
Max altitude: 110 km
Microgravity: Three minutes
Passengers/Pilots: One/one
Status: 2012: expected test of mark I prototype. 2013–14: expected flight of mark II production model.



4 NEW SHEPARD

Company: Blue Origin
Max altitude: 100 km
Microgravity: Three minutes
Passengers/Pilots: Three/none
Status: 2006–11: at least three test flights. 2011: NASA funds crew-capsule development.

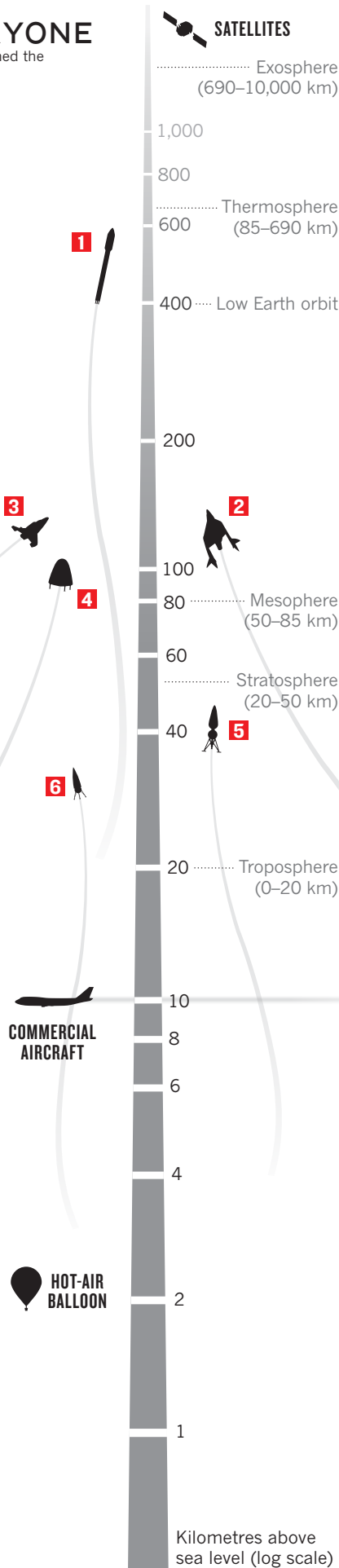


5 SUPER MOD

Company: Armadillo Aerospace
Max altitude: 40 km (initial flights)
Microgravity: Three minutes
Passengers/Pilots: Unmanned
Status: 2010–11: many test flights. 2011–12: expected flight of new commercial vehicle to 40–100 km.

6 XAERO

Company: Masten Space Systems
Max altitude: 30 km (initial flights)
Microgravity: Three minutes
Passengers/Pilots: Unmanned
Status: 2011: expected test flight to 30 kilometres. 2012: expected flight of larger Xogdor to 100 km.



has relied on private rockets for decades. Most communications satellites and even classified military payloads are sent into orbit atop commercial launchers. Similarly, unmanned private sub-orbital flights are routine, with various companies marketing 'sounding rockets' that take measurements and perform experiments in space.

The problem is that these launchers are so expensive that only the government and large telecommunications corporations can afford them. One reason for their cost is that they are expendable, discarded after a single use. In the 1970s, NASA tried to eliminate much of that waste by developing the fleet of space shuttles, which were partially reusable. But in practice, the advantage of reusability was more than offset by the difficulty of engineering a vehicle that could withstand the stresses of launch and re-entry, and be ready to fly again; the shuttles proved to be hideously pricey.

Now, with the shuttles nearing retirement, NASA has been trying again to get the launch costs down — this time by encouraging the private sector to develop cheaper rockets and crew capsules to reach low Earth orbit. Last month, for example, the agency awarded \$269 million in development money to four companies, one of which — the SpaceX Corporation of Hawthorne, California — has already had a successful launch with the Falcon 9 rocket, which first reached orbit in June 2010.

COMMERCIAL ENTERPRISE

For the time being, however, more entrepreneurial energy is focused on the suborbital regime, in which the costs are lower and potential customers are more plentiful. The suborbital race began in 2004, when test pilot Mike Melvill repeatedly flew a privately developed reusable spacecraft, *SpaceShipOne*, to altitudes of more than 100 kilometres in the skies over Mojave, California. Shortly afterwards, Burt Rutan, the craft's designer, partnered with entrepreneur Richard Branson to form Virgin Galactic, a venture to fly tourists on \$200,000-per-seat, 110-kilometre-high suborbital jaunts using a fleet of 6-passenger 'SpaceShipTwo' spaceplanes, which are currently in development.

A number of other start-up firms followed — many with roots in the computer and Internet industries, a testament to the symbiosis between space dreams and lucrative high-tech careers. The companies include Armadillo Aerospace, organized and funded in Heath, Texas, by John Carmack, the computer-graphics wizard behind the hit videogames *Doom* and *Quake*. There is also Blue Origin of Kent, Washington, founded by Jeff Bezos using a fraction of the fortune he earned creating Amazon.com; Masten Space Systems of Mojave, established by David Masten, a former information-technology networking guru; and XCOR Aerospace, also of Mojave, headed by Jeff Greason, an engineer who helped to develop the technology used in Intel's Pentium line of computer chips (see 'Space for everyone').

These entrepreneurs expect one of the most lucrative applications for suborbital spaceflight to be space tourism, but tourist flights won't begin for at least a year, and probably two. In the meantime, to flesh out launch manifests and help to subsidize unmanned test flights, companies have begun courting research institutions, government agencies and independently wealthy investigators who want to run scientific experiments in suborbital space. All five companies sent representatives to the NSRC this year, hoping to court more clients like the SwRI, which has bought a total of eight seats, and options for nine more, on suborbital flights, split between Virgin Galactic's *SpaceShipTwo* and XCOR's *Lynx*.

Few of the attendees in Orlando needed much convincing; just about everyone there seemed eager to climb on board. Astronomers talked about raising telescopes above the atmosphere to perform mid-infrared searches for water on the Moon and to observe planets, comets and asteroids. Planetary scientists detailed microgravity experiments to investigate the collisions of dust and sand that



Daniel Durda (left) and Alan Stern (floating) train for zero-gravity on an aircraft.

form the building blocks of planets. Atmospheric scientists discussed *in situ* sampling of the 'ignosphere', the largely unexplored stratum of the upper atmosphere that lies above the altitudes attainable by weather balloons, but below those of satellites. Materials scientists were eager to study how microgravity affects processes such as combustion.

But the enthusiasm of individual researchers is one thing. Getting the institutions they work for to pay for tickets is something else. NASA, for example, has signed contracts with Armadillo and Masten, paying nearly \$500,000 for seven flights carrying engineering equipment. But the agency does not yet have approval to buy seats for suborbital passengers — both Armadillo and Masten are currently focusing more on unmanned flight — and none of the flights it has purchased will actually reach space. The highest planned altitude is about 40 kilometres.

To sell those passenger seats, the launch companies will have to convince decision-makers that their reusable vehicles offer significant advantages over existing ways to access the weightless space environment. For only a few tens of thousands of dollars per trip, researchers can book custom-modified aeroplanes that fly in a series of parabolic trajectories, providing microgravity in 30-second bursts for a total of 5–10 minutes per flight. Or, for an admittedly steep \$1 million to \$2 million, researchers can put automated equipment on a sounding rocket that provides up to 20 minutes of microgravity far above Earth's atmosphere.

The new commercial vehicles vary in their capabilities, but generally fall between the two existing options: they can reach between 30 and about 100 kilometres in altitude and offer 3–5 minutes of weightlessness.

Stern and other proponents believe that the reusable space vehicles' short times in space are counterbalanced by their high frequency of flights. "As a principal investigator, it took me nearly a decade to get seven flights on NASA's sounding rockets, but [the SwRI is] going to be flying these eight missions in the space of about a year," says Stern. "Virgin Galactic alone will be flying daily with six vehicles; XCOR is going to fly four times per day; Blue Origin says it'll fly once a week. This will give us unprecedented access to space."

THE HUMAN ELEMENT

But do humans really need to ride along on such experiments, with all the risks and complexity that entails? "The advantage is twofold," says Stern, making much the same argument that human-spaceflight proponents have been making since NASA's Apollo programme of the 1960s. "First, you don't have to pay to automate your experiment any more. And second, that means you can easily react to your data collection in real time and make changes in your experiment to get better results."

To demonstrate those advantages, the SwRI researchers are planning three showcase experiments. One, 'Box of Rocks', is a transparent case of stone fragments and ceramic bricks meant to simulate how

loose material settles on asteroids with low surface gravity. Another uses a refurbished ultraviolet telescope, which flew on a space shuttle in 1997, to observe astronomical objects and upper-atmosphere phenomena through a cabin window. The last uses a 'bioharness' to monitor and record how the blood pressure, heart rate and other physiological parameters of passengers vary under the flight profiles of different vehicles.

Paul Hertz, the chief scientist in NASA's Science Mission Directorate, remains sceptical. He is broadly optimistic about the science potential of reusable suborbital vehicles, but less convinced that involving humans in experiments requiring astronomical observations will be useful. "Because these flights are relatively short, it's really not that constraining to pre-program your observing plan," he says.

The weight of seats and life-support exacts a huge performance penalty, agrees Stephan McCandless, an astronomer at Johns Hopkins University in Baltimore, Maryland. McCandless has flown ultraviolet astronomy experiments using sounding rockets for more than two decades, and what the new reusables can do "pales in comparison", he says.

"Sounding rockets are extremely valuable," Stern concedes. "They go higher, have more sophisticated pointing systems, and can carry payloads commercial reusables just can't. But reusables can fly far more frequently, they are 10–20 times lower in cost, and they can bring scientists along with their instruments. This is a debate about having a fork or a spoon at the table — they are for different purposes."

John Grunsfeld, a former shuttle astronaut and current deputy director of the Space Telescope Science Institute in Baltimore, is also doubtful that the new vehicles will be better than sounding rockets, but admits that reusables could offer fresh opportunities for science.

"The real potential to benefit science is not necessarily in this first wave of vehicles, but in their prospect for a future in which crewed commercial vehicles will routinely and frequently access longer periods of weightlessness, or better yet, reach Earth orbit," says Grunsfeld. That ready availability, in turn, could facilitate imaginative science that might not make it through the peer-review processes of government agencies and major academic institutions. Of course, cautions Grunsfeld, "the value of that science remains to be seen".

For Stern, the value of routine suborbital spaceflight shouldn't be measured only in grants awarded and peer-reviewed papers published. Speaking at one NSRC session, he told the audience that they should feel no shame if their interest in suborbital science stemmed mainly from their yearning to fly in space.

"I don't think most scientists appreciate very well how motivational human spaceflight can be," Stern said. "Going into science is hard — there are easier careers where you can make more money. But when everyday educators and working researchers can visit classrooms and speak to schoolchildren about personally going into space, that has real effects. We can contribute to the future by giving birth to this new industry and the opportunities it brings."

That message seemed to resonate at the NSRC. Between and after the sessions, discussions about suborbital science frequently segued into conversations about inspiring the next generation of researchers to do great things — to beam energy to Earth from vast solar panels on satellites, to visit asteroids, to colonize the Moon or to travel to Mars. Indeed, for most of the scientists gathered in the conference rooms and hotel bars, the prospect of democratized suborbital flight seemed to be a blank screen on which they could project their current plans and future dreams of humanity's expansion into, at minimum, the rest of the Solar System.

"A lot of people are reluctant to talk about the big picture, and they may not be able to always clearly articulate why they want to take these trips, but what many of them want is to take part in making the future happen," says Greason, XCOR's chief executive. He is no exception. "The reason why I'm in this business is because I think it has the potential to be the beginning of something that will last for a very, very long time." ■

Lee Billings is a freelance writer in New York.

COMMENT

ENVIRONMENT The age of cheap, abundant water is ending **p.27**

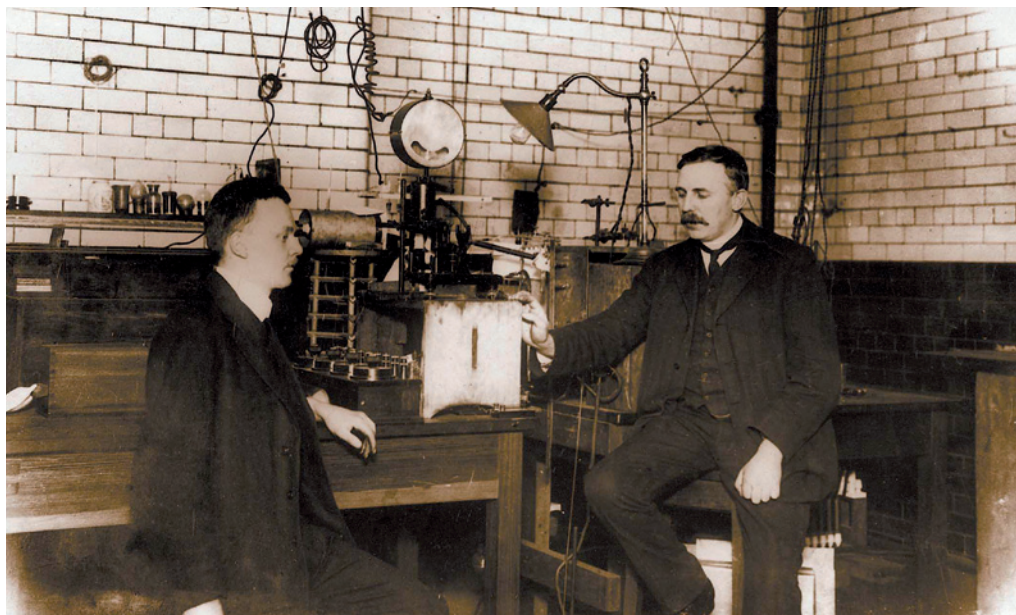


FILM Werner Herzog on cave painting and the brutality of nature **p.30**

POLICY Overhauling the NIH will not fix the broken US health pipeline **p.31**

COMMUNICATION Public mistrust of Japanese officials is not irrational but sensible **p.31**

D. TUFFS/THE GUARDIAN



Hans Geiger (left) and Ernest Rutherford's experimental work revealed the nucleus at the centre of atoms.

From isotopes to the stars

Creating more exotic isotopes will reveal the stellar formation of atoms — a fitting tribute to Ernest Rutherford, say **Michael Thoennessen** and **Bradley Sherrill**.

A hundred years have passed since Ernest Rutherford published his discovery of the atomic nucleus in May 1911 and started a journey to the centre of the atomic world. In Rutherford's famous experiment, a stream of α -particles was aimed at a very thin sheet of gold foil. Some of the particles were deflected at angles that suggested they had collided with a small, dense atomic core. As Rutherford remarked: "It was almost as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you." The experiment supported a planetary model of the atom — the idea that most of the mass is concentrated in a nucleus, with even smaller electrons orbiting it like planets around the Sun.

Over the century since, scientists have probed to sizes 1,000 times smaller than Rutherford managed — to the level at which quarks are important — and developed the

standard model to describe these particles and their interactions. This quest to find nature's ultimate building blocks is being led by experiments at CERN, Europe's particle-physics laboratory near Geneva, Switzerland.

Despite this progress, some basic questions remain unanswered. It is not known how Rutherford's nucleus can result from quarks and the strong force. It is not even known in detail how the strong force binds quarks to make neutrons and protons, or how it results in the forces that hold together protons and neutrons in the nucleus. Even simpler questions, such as how many elements might be possible, or how many neutrons a given number of protons can bind, are currently unanswerable.

So, away from the high-profile, high-energy frontier, a small army of machines is quietly advancing understanding of the atomic nucleus by generating new and rare

nuclides — atoms with a specific number of protons and neutrons in their nuclei. In 2010, for the first time, more than 100 new unstable isotopes — nuclides with different numbers of neutrons — were discovered in a single year (see "The nuclide trail"). We expect more than 1,000 new isotopes, including some of the most scientifically interesting to date, to be discovered over the next decade or so.

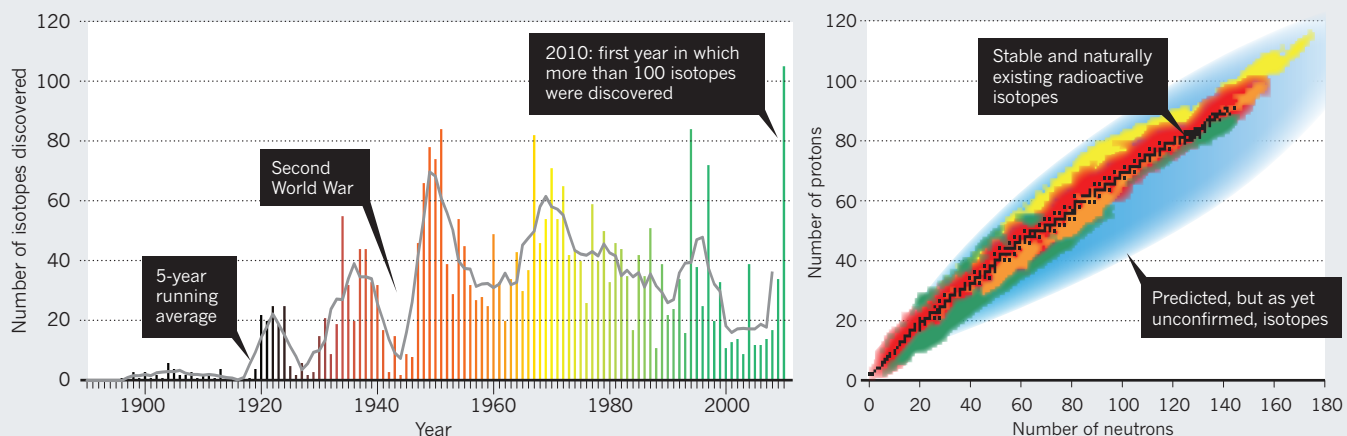
Initially, the search for new isotopes was driven by the quest for the unknown, to make something nobody else had made and the urge to understand the underlying forces. But they have enormous practical applications too: in nuclear energy, medical imaging and treatment, carbon dating and tracer elements. The international effort to create new isotopes will push our understanding of atom formation and nuclei to new levels, but may also lead to applications.

Even before Rutherford's experiment, ►

UNIV. MANCHESTER

THE NUCLIDE TRAIL

Isotope discovery over the past 100 years (below) has jumped with each introduction of new technology. Some 2,700 radioactive isotopes have been discovered so far (below right), but about 3,000 more are predicted to exist.



▶ radioactive-decay studies showed that a given element can exist in different forms. The discovery of the neutron in 1932 revealed that the nucleus of an atom was composed of protons and neutrons. Soon after, Irène Curie and Frédéric Joliot used α -particles from polonium and targets of boron, magnesium and aluminium to create the first radioactive isotopes in the laboratory. The new isotopes of nitrogen, aluminium and phosphorus had one neutron fewer than the normal stable nuclides of these elements.

Since then, researchers have been searching for the limits of nuclear existence, to discover what element may have the most protons and what are the largest (and smallest) number of neutrons for a given element. Even today, the limit to the number of neutrons that an element can bind is known only for the lightest elements, from hydrogen to oxygen. That is one very small corner of the possible nuclear landscape (see 'The nuclide trail').

There are almost 300 stable nuclides on Earth and another 2,700 radioactive isotopes have been identified so far. This represents perhaps only half of all predicted isotopes. Around 3,000 have yet to be discovered (it might be as many as 5,000 or as few as 2,000). Although the different masses of isotopes do not influence their chemistry much, the production and study of rare isotopes is crucial to understanding the process of nature that makes atoms in their birthplace.

Most of the elements in nature are created in stars and stellar explosions, and the isotopes involved are often at the very limits of stability. The next generation of rare-isotope accelerators will create, for the first time on Earth, most of the isotopes that are formed in stellar environments. Where physicists currently have to rely on theoretical models based on extrapolations, they will soon measure the properties of most of these isotopes directly. It could help to answer

other fundamental astrophysics questions on where in the cosmos these isotopes are created, why stars explode, the nature of neutron stars and what the first stars in the Universe were like.

MARCH OF MACHINES

The first particle accelerators, developed in the early 1930s, revealed many new isotopes. The Second World War delayed progress but, afterwards, neutron-capture and neutron-fission reactions in nuclear reactors continued the exploration. The next advance was the development of heavy-ion accelerators in the 1960s, which produced heavy neutron-deficient isotopes in fusion evaporation reactions.

With higher-energy accelerators in the 1990s, scientists could create more neutron-rich nuclei during in-flight fission or projectile fragmentation of high-energy heavy ions. This has been the most productive route to isotope discovery in recent times. But in the past decade, the rate of discovery dropped to levels not seen since the 1940s. It became obvious that dedicated rare-isotope accelerators were needed to make further progress.

The first of these facilities, the Rare Isotope Beam Factory, came online in 2007 in Wako, Japan. In 2010, it reported the discovery of 45 new neutron-rich isotopes.

To ensure that this is the beginning of a new era rather than just a discovery spike, it is crucial to continue efforts worldwide. Centres are under development, such as the Facility for Antiproton and Ions Research in Darmstadt, Germany, SPIRAL2 in Caen, France, and the Facility for Rare Isotope Beams at Michigan State University in East Lansing. Scientists in the United States have been trying to build a rare-isotope accelerator for almost 20 years. Funding for an earlier facility was halted during a previous period of austerity.

Today's difficult financial conditions must

not be allowed to halt the machine builders. The facility in Germany still needs to secure sufficient funding to start operations by the end of the decade. Isotope discovery over the past 100 years has been a worldwide effort, with more than 3,000 scientists in 125 laboratories in 27 countries contributing. It will be a shame if the German facility — an international collaboration from its outset — does not move forward expeditiously.

Pushing science to the limits produces surprises. We have already learned that rare nuclei with extreme proton-to-neutron ratios don't always follow the textbook behaviour of known stable isotopes. For example, the size of stable nuclei is proportional to their mass — it scales as $A^{1/3}$ (where A is the mass number of neutrons and protons). However, this simple relationship ignores any differences between neutrons and protons. Some rare nuclei that exist only fleetingly have proved to be significantly larger.

Other surprises may be in store. Hopefully, the next-generation facilities will create more than 1,000 new isotopes, and the limit of nuclear existence will be pushed towards heavier elements, up to zirconium (40 protons) but still some way from gold (79 protons). Fundamental phenomena are waiting to be discovered, and increased production of rare isotopes will bring new applications in medicine and other fields. We are confident that in the next 10–15 years, most of the isotopes needed to answer the question 'What is the origin of elements in the cosmos?' will be created in the lab for the first time. A fitting tribute to Rutherford. ■

Michael Thoennessen and Bradley Sherrill are at the National Superconducting Cyclotron Laboratory and the Department of Physics and Astronomy at Michigan State University, East Lansing, Michigan 48824, USA. e-mails: thoennessen@nscl.msu.edu; sherrill@frib.msu.edu



In some countries, two in every five women still face a long daily walk to collect water.

ENVIRONMENT

Water, water everywhere...

Margaret Catley-Carlson wonders why humanity places so little value on its most basic resource.

As supplies around the world come under pressure, is it all over for water? In this comprehensive, entertaining and torrential flow of a book, journalist Charles Fishman answers with a definitive no. But "the golden age of abundant, cheap and safe water" is quickly disappearing. We needn't panic, he says, but it isn't going to be like it was. We must treat water differently. And many of us are in for a rude shock.

The Big Thirst is a key read for people who wonder how water became so scarce that in

2007–08 the cities of Atlanta, Georgia, and Barcelona, Spain, almost ran out, and why in some countries around the world two out of five women still walk long distances each day to collect water. Informative and wide-ranging, it covers how water molecules were formed in interstellar clouds and came to Earth after the Big Bang; how ultra-pure water used in micro-chip manufacture is so clean that it is toxic to

human touch; and the probable existence of several oceans' worth of water sequestered in rocks hundreds of kilometres below ground.

Discussions of global water management often drown readers with mega numbers. Fishman asks instead: what do billions of gallons or cubic metres, or trillions of investment deficits in water infrastructure, look like or signify? He makes

lively comparisons: every day, the United States flushes more water down its toilets than either Canada or the United Kingdom consumes in total; and one shipload of water delivered to drought-stricken Barcelona in 2008 supplied the city for just 32 minutes.

Newcomers to the water issue are usually relieved to find out that there remains enough water on the planet to supply the needs of humans and the ecological system; the question is how to manage it. But answering that question is not simple. Fishman places the responsibility for difficulties in water management firmly where it belongs — on the witch's brew of sociology, economics, suspicion, electoral politics, history and mythology that makes decision-making sometimes difficult, and often nearly impossible.

The number-one problem is that water is not valued. In our lives, businesses and habits we abuse and overuse this multipurpose solvent, precious elixir and indispensable substance. Transportation and energy projects, fresh agricultural developments, new suburbs and shopping malls are embarked on without thought about their effects on local water. Will it be polluted? Is ground water running out? Who is downstream and what will be the impact on them?

The second problem is that because we don't value water, we are reluctant to pay for it, or for the reservoirs, pipes, energy, chemicals, staff, fencing and monitoring needed to get clean water to the point of use. So municipal pipes leak, and many cities across the world lose from one-quarter to one-half of the water in their plumbing systems. Most irrigation systems are less than 50% efficient.

These issues combine to create a political 'no-go' zone. Politicians will lose elections if they vow to charge more for water, and those who favour new development routinely win. Fishman paints dramatic pictures of the results. In some Indian cities, water is available for only two hours every two days, so each household must set up its own water-storage facilities. Residents of Atlanta continued their wasteful water-use habits as the reservoirs



The Big Thirst: The Secret Life and Turbulent Future of Water

CHARLES FISHMAN
Free Press: 2011.
400 pp.
£16.67/\$26.99

NATURE.COM
For more on water in
the Middle East:
go.nature.com/qgekxz

dropped and elected officials prayed for rain. But even when the rains come, the problem is not fixed; and drought is sure to recur.

Fishman enjoys naming and shaming the villains. But he takes greater joy in celebrating the heroes: the laundries in Las Vegas, Nevada, and the citizens of Australia's Gold Coast who now recycle urban water. Fishman explores at length the paradox that whereas companies such as Coca-Cola — headquartered in Atlanta — and Campbell Soup, of Camden, New Jersey, have set themselves elaborate water strategies and water-saving measures, most cities, including Atlanta, have not.

He both praises and damns the private sector. The market can drive efficiency savings, he says, but it also creates solutions for problems that don't exist by, for example, "foisting bottled water on a too-gullible world", and fails to fix the real problems. He sees little future for a trade in water, because water cannot be transported easily over long distances. It's costly, politically and practically. Yet 'virtual' water — used in the production of coffee, T-shirts, cars and everything else we make — is traded with little heed for its economic or ecological value.

Technological advancement is and will be important, and Fishman covers it nicely. Given that the agricultural sector uses more than 70% of the global water supplies, surely everyone would be cheered by the idea of a high-yielding new crop variety that can mature using only 40% of the water? But if those crops are genetically engineered, more than one continent will recoil. Farmers who get water free or for little cost have no incentive to reduce their usage with water-saving devices. Nor are many of the new technologies taken up, even though someone invents a water purifier nearly every week that 'for only pennies per day will provide a family with clean drinking water'. The dispiriting truth is that few are bought.

The Big Thirst is a delight to read — full of salient and fascinating examples, well-researched and laced with wry humour. It would be wonderful if Fishman's rant against bottled water converted every reader. It would be even better if it promoted a serious reflection on how little we value that on which our life depends. ■

Margaret Catley-Carlson serves on water and agricultural boards and advisory committees, including the World Economic Forum's Global Agenda Council on Water Security and the United Nations Secretary General Advisory Board on Water.
e-mail: m.catley-carlson@cgiar.org



A blogger (left) for an Internet radio station in Egypt that fights intolerance towards divorced women.

TECHNOLOGY

Together, bit by bit

A historian's insights into digital culture fascinate George Rousseau.

Respected intellectual historian Milad Doueihi describes himself as an "accidental digitician" — by his own admission more a user of information technology than a creator of it. Such people, he argues in *Digital Cultures*, are forging a new global culture. The impact of computers on our minds, bodies and societies is already far-reaching. Whether we like it or not, digital culture is permanently entrenched.

Douihi, an expert on literacy, points out that the voices of historians have largely been missing from discussions of the Internet. By showing how modes of communication and human relationships have changed since its rise, he makes a persuasive case that digital culture has broken free from print culture, which extends from the Gutenberg Bible of the 1450s to the present. Instant response, brevity, minimal spelling and grammar, novel syntax and different modes of composition have created new forms of literacy.

As a consequence, the way we view our identity, citizenship and political selfhood has changed. Doueihi sees blogging as "one of the greatest success stories". With the rise of online forums, everyone can communicate freely without publishers' intervention. As a result, we are more dedicated to the Internet than to any other civic cause, or even to our everyday work. As well as rich and poor, there is now another great social divide: between those with and without access to these web conversations.

In our online interactions, a new civility has emerged, along with the uncivilized behaviour — 'trollism' — that results from online anonymity. Urban dwellers blog more than those outside cities, and have created parallel cities in the blogosphere. And podcasts have reinvigorated the voice.

Douihi's argument for a culture shift rests on three components of the online world. One is its creation of an anthology. The digital culture, rather than creating long, sustained narratives, assembles fragments of material — but not into logical wholes. We invest everything in e-mail responses rather than saving up our thoughts for long letters or books. All these snippets can then be assembled by different readers in different ways.

Douihi also briefly cites religion as a central aspect of any new culture, although he never explains what he means by the word 'religion'. He passes quickly on to the third component — group identity, arguing that we seem to have a greater craving for belonging than previous generations.

Digital group identity, says Doueihi, differs from previous print-based concepts in several



Digital Cultures
MILAD DOUEIHI
Harvard Univ. Press:
2011. 175 pp.
£14.95/\$19.95

ways: speed of communication, multiple numbers of readers instantly reached, and the assumption that everyone who receives your digital message is interested in what you say. But it has a downside. Someone who paid two shillings for a book in the eighteenth century worked a week to buy that book and wanted to own it. With so much to choose from, readers of blogs may never find an account of such value to them.

The new types of 'group belonging' arising on the Internet, through which people achieve personal popularity and find safety, are creating a new emotional comfort zone. This begs for a broader discussion of emotional, moral and other types of literacy, which Doueihhi does not address. I also craved more knowledge about the interior world, especially the affective and emotional resonances of web users, many of whom are young.

Doueihhi has sensitive antennae for the legal ramifications of the new digital culture, as his debates on intellectual property rights, security and related issues show; and he may be right that at the root of these controversies is the annihilation of the old conception of what it is to be an author. In the print culture, the author controls the material that is read; in the new culture the reader is empowered to contribute, as in the shared editing of Wikipedia.

Many historians will counter that aspects of print culture — such as sustained narrative and religions organized by ethnic and national identity — are not defunct. We may spend our time in global digital cities, but our passports are not yet shredded. Doueihhi might reply that this is a matter of degree: some civic forms have changed more rapidly than others. Our expectation of what a book is remains the same.

Although Doueihhi bypasses the scientific community as a specific case, the new digital literacy must have altered what it means to be a scientist, especially in terms of identity and group belonging. Celebrity culture among scientists has undeniably become more frenzied in recent decades. Yet the effect of the Internet on the process of doing science is more elusive. With thousands of electronic messages traversing a typical laboratory each day, it will be increasingly difficult for sociologists to disentangle how networks of people manufacture scientific facts, in comparison with earlier accounts such as Bruno Latour and Steve Woolgar's *Laboratory Life* (1979).

Written in the 'old' discursive format, *Digital Cultures* includes much to think about. The pace of change is fast, but Doueihhi's insight is fresh. ■

George Rousseau is a professor of history at the University of Oxford, co-director of its Centre for the History of Childhood, Oxford OX1 4AU, UK, and author of *Nervous Acts: Essays on Literature, Culture and Sensibility*. e-mail: george.rousseau@magd.ox.ac.uk

Books in brief



Born in Africa: The Quest for the Origins of Human Life

Martin Meredith SIMON & SCHUSTER 432 pp. £16.99 (2011)

More than a century after Charles Darwin suggested that the ancestors of modern humans might lie buried in the African plains, we are still piecing together the jigsaw of our evolutionary past. Journalist and historian Martin Meredith tells the story of the palaeontologists who sought the bones of early hominids there, from the discovery of skeletons in Tanzania's Olduvai gorge in the early twentieth century to the latest genetic research on the branches of the human family tree.



Rising Force: The Magic of Magnetic Levitation

James D. Livingston HARVARD UNIV. PRESS 288 pp.

£20.95 (2011)

Giving a new meaning to literary suspense, physicist James Livingston devotes his book to the science of magnetic levitation. From laboratory demonstrations of floating magnets, flying frogs and suspended sumo wrestlers to the realities of urban maglev trains, he uncovers humanity's fascination with the magic of defying gravity, as well as the physics of magnetic fields and superconductivity.

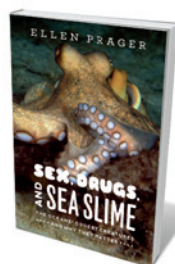


Divine Machines: Leibniz and the Sciences of Life

Justin E. H. Smith PRINCETON UNIV. PRESS 392 pp.

\$45/£30.95 (2011)

Seventeenth-century philosopher G. W. Leibniz is best known for his mathematical discoveries, including calculus. But he also investigated the science of life. Philosopher Justin Smith describes how Leibniz's experimentation in medicine, physiology, taxonomy and palaeontology influenced his philosophical ideas, causing him to shy away from mechanical views of nature towards more organic ones.

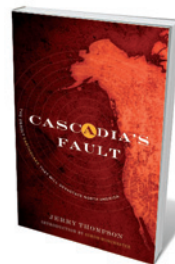


Sex, Drugs, and Sea Slime: The Oceans' Oddest Creatures and Why They Matter

Ellen J. Prager UNIV. OF CHICAGO PRESS 216 pp.

\$26/£17 (2011)

Beneath the waves, anything goes, explains marine scientist Ellen Prager in her tour of some of the saltier habits of sea life. From the inside-out posture and bioluminescent fireworks of the vampire squid to the mucus deluge that protects the slimy hagfish, she explains how marine critters adopt unusual approaches to sex, predation and defence. And she explores how these diverse creatures, from krill to the grey whale, are crucial for our food supply, economies and even drug discovery.



Cascadia's Fault: The Earthquake and Tsunami That Could Devastate North America

Jerry Thompson and Simon Winchester COUNTERPOINT PRESS

352 pp. £16.06/£26 (2011)

Following the recent devastation in Japan, journalist Jerry Thompson points out with unfortunate timeliness that North America is also at risk from a cataclysmic earthquake and tsunami. The Cascadia subduction zone stretches 800 kilometres from Vancouver Island to northern California, where the ocean floor slips below the continent. He follows the researchers who monitor the area, and asks what would happen if a magnitude-9 quake and 30-metre waves hit Vancouver and Seattle, Washington.



Werner Herzog charts the emergence of a new human sensibility 35 millennia ago in his latest film.

Q&A Werner Herzog Illuminating the dark

As he releases a 3D documentary about the prehistoric paintings in Chauvet Cave in southern France, Werner Herzog — the German director of *Fitzcarraldo* and *Grizzly Man* — talks about cave art and the hostility of nature.

What drew you to cave art?

It dates back to my adolescence. I come from a remote mountain valley where we had no telephone, no radio, no running water. A book in a bookstore caught my attention. I was mesmerized by a prehistoric picture of a horse — perhaps from Lascaux Cave. I was always interested in archaeology because of my grandfather, an archaeologist who did his life's work on a Greek island close to the Turkish coast. He excavated a huge site that includes temples and a medical spa where ancient doctors would work. When the chance came to film in Chauvet Cave I was immediately on board.

Why is Chauvet special?

Some of the most wonderful caves with prehistoric art, such as Lascaux in the Dordogne in France and Altamira in the Spanish Pyrenees, have had to shut because of problems with mould. Chauvet, in the Ardèche in France, was preserved as the perfect time capsule. Owing to the collapse of the face of the gorge, the cave entrance was sealed for roughly 20,000 years. And when the cave was discovered in 1994, the explorers did everything right. They rolled out plastic sheets and crawled along them to avoid stepping on the floor. They found the tracks of cave bears,

which had been extinct for tens of thousands of years. And the charcoal remains of fires made to illuminate the paintings. One swipe mark of a torch on the wall was radiocarbon dated to nearly 30,000 years ago. The paintings themselves date from 30,000 to 35,000 years ago.

What do the paintings show?

The bestiary is limited and mysterious. The animals depicted range from reindeer to woolly mammoths, woolly rhino, lions, bison — huge, dangerous, powerful beasts, and not only animals that you would hunt. There is no sign of a fox, weasel or bird, except one scratched image of an owl. Painting never got any better through the ages, not in ancient Greek and Roman antiquity, nor during the Renaissance. It's not like the *Flintstones* — the work of crude men carrying clubs. This is the modern human soul emerging vigorously, almost in an explosive event. You sense the presence of the artists because it's so fresh: we felt that eyes were looking at us from the dark.

What do we know about the cave artists?

For the time, they were high-tech. An ice-free corridor would have connected Chauvet to the Swabian Alb, 400 kilometres away in southern Germany, where flint tools and

bone and ivory flutes have been found. The cave was never inhabited, although there were burials in the region. Strangely, Chauvet people only painted deep inside the cave, where it was completely dark. Some archaeologists claim the pictures have ritualistic or shamanistic meanings. But we simply do not know.

The Cave of Forgotten Dreams (3D)

DIRECTED BY
WERNER HERZOG

Now showing at
US/UK cinemas

What were the filming challenges?

We were allowed one week of shooting, but just four hours per day. We had to move along a metal walkway. No more than three camera people, sound or 3D specialists could assist me, and we had to use lightweight equipment that did not emit any heat. It was tough: 3D apparatus is large and clumsy, and must be reconfigured for each type of shot. When the camera moves closer to an object, the lenses have to move towards each other and 'squint'. We had to do these high-tech things in semi-darkness with only a few torches.

Was filming in 3D worth the trouble?

It was. The formation of the cave is very dramatic. There are bulges and niches and pendants, which the artists also utilized in their drama. For example, a huge bulge in the rock now is the bulging neck of a charging bison; a horse comes out shyly from the recesses of a niche. When you see the film you know immediately that it was the right thing to do. Otherwise, I'm sceptical of 3D.

You've said you see nature as hostile and chaotic. Why?

I've heard too many times that there's a cosmic harmony. This vapid new-age babble enrages me. The Universe is not harmonious and beautiful, it is dangerous and hostile. My opinion is evident in *Grizzly Man*, for example, which is about a man who went out to Alaska to protect grizzly bears by standing a couple of metres away from them. Even our supposedly benign Sun is a danger — hundreds of thousands of simultaneous atomic explosions. Imagine how destructive a black hole would be. Yet the more we know, the more fascinating it gets. There's an inherent curiosity in the human race to understand the Universe that's around us. That distinguishes us from the cow in the field. ■

INTERVIEW BY JASCHA HOFFMAN

CORRECTION

The Books in Brief summary of *The Sorcerer's Apprentices* by Lisa Abend (*Nature* 471, 577; 2011) wrongly suggested that she underwent training as a chef; in fact, she observed training.

CORRESPONDENCE

NIH revamp: US health care at fault

Contrary to Michael Crow's implications (*Nature* **471**, 569–571; 2011), the annual budget of the US National Institutes of Health (NIH) of US\$31 billion is only a small percentage (barely 1%) of yearly US health spending — now \$2.5 trillion.

US health care is costly because it is the only wealthy industrialized country without public health insurance. Its citizens give vast sums to insurance companies whose primary function is to bleed money from the system while maximizing profit.

Everyone should benefit when an NIH-funded discovery is made that extends human life. But a shamefully large fraction of the US population does not because they have inadequate health insurance, if any. More politically enlightened nations stand to gain more by providing the best possible health care.

Thomas E. DeCoursey *Rush University Medical Center, Chicago, Illinois, USA.*
tdecours@rush.edu

NIH revamp: real issue is resources

A failure to translate the United States' global leadership in biomedical science into a comparable position in health care (*Nature* **471**, 569–571; 2011) does not justify dismantling the very source of that leadership — the National Institutes of Health (NIH). The real issue is that more science, data and resources are needed by other units of the US Department of Health and Human Services (HHS) responsible for engineering the application of discoveries.

Opportunities for scientific reorganization in the NIH include improving cost-effectiveness and

instrumentation of assets and weaknesses. But it is crucial to separate the engine of discovery from the engine of application. Discovery is stochastic and opportunistic; application is the stuff of engineers. That is why attempts to over-engineer discovery fail and why science should not drive its application.

There should be separate units to promote discovery, assess outcomes and engineer the healthcare system. At present, these approximate to the NIH, the Agency for Healthcare Research and Quality, and HHS units such as the Food and Drug Administration.

Russ Altman *Stanford University, Stanford, California, USA.*
russ.altman@stanford.edu

NIH revamp: avoid a redundant revolution

A restructuring of the US National Institutes of Health (NIH) to include new institutes for “health transformation” and for research into “health outcomes”, as Michael Crow advocates (*Nature* **471**, 569–571; 2011), is unnecessary. These would duplicate the function of agencies that, like the NIH, are already overseen by the Department of Health and Human Services.

The Agency for Healthcare Research and Quality focuses on outcomes research. The National Coordinator on Health Information Technology and the Centers for Medicare & Medicaid Services (CMS) focus on transformation. In particular, the CMS will administer US\$10 billion from the 2010 Affordable Care Act for research related to sustainable cost models for health care (<http://innovations.cms.gov>).

John Robinson *South Dakota State University, Brookings, South Dakota, USA.*
john.robinson@sdsu.edu

UNESCO helps manage tsunamis

In disasters on the scale of Japan's 11 March tsunami, every second counts in making accurate information available to those who need it most. To this end, the United Nations Educational, Scientific and Cultural Organization (UNESCO) helps professionals and populations to anticipate the risks, assess possible flooding and coordinate monitoring.

Some lessons have been learned from the ravages of the 2004 tsunami in the Indian Ocean. In addition to the Pacific early-warning systems, UNESCO's Intergovernmental Oceanographic Commission is coordinating the set-up of regional tsunami-warning centres in the Indian Ocean, the Caribbean, the north-east Atlantic and the Mediterranean, as well as full-scale simulation exercises.

International scientific cooperation can help in countering such disasters, whose scope extends beyond frontiers and state capacity. But this cannot replace the authority and initiative of national leaders. We also need to do much more to strengthen the capabilities of local communities.

Managing the unexpected depends on education and culture. For example, Japanese children are taught how to respond to earthquakes and tsunamis at school; and because the people of Simeulue Island in Indonesia were aware of tsunami warning signs, only seven died in the 2004 event. With UNESCO's support, Indonesia and Thailand are accelerating their risk-reduction education. Last year, students and teachers were trained in schools across six coastal cities in Colombia, Ecuador, Peru and Chile.

Urbanization and uncontrolled development

threaten the coral reef and mangrove ecosystems that mitigate the force of tsunamis. As some 10% of the global population live in low-lying coastal zones, protecting these natural barriers is also a shared responsibility.

Irina Bokova *Director-general of UNESCO, Paris, France.*
dg@unesco.org

It is rational to doubt Fukushima reports

Officials have no right to dismiss as “irrational” the public's mistrust of official pronouncements about the 11 March earthquake and tsunami damage to Japan's nuclear reactors in Fukushima.

The public in Japan and elsewhere has figured out two things about Fukushima. First, what might happen next is a potentially bigger problem than what has happened so far; and second, governments, experts and authorities have been consistently behind the curve in talking openly about what might happen next.

People are suspicious of official assurances that the current situation will get no worse, maybe rightly. They don't trust the authorities to tell them the ways in which it could get worse and how likely it is to do so. Many don't even trust the authorities to tell them promptly if it does.

As a result, a variety of precautions that might be considered excessive or premature if the public felt they could trust the authorities — avoiding Japanese foods, for example, or seeking out a supply of potassium iodide — suddenly become sensible and should not be branded as illogical, hysterical or radiophobic.

Peter M. Sandman, Jody Lanard *Risk Communication Consultants, Princeton, New Jersey, USA.*
peter@psandman.com

The origins of novelty

Treehoppers produce highly diverse structures called helmets. To do so they seem to have exploited the genetic potential, long inhibited in other winged insects, to develop wings on a particular anatomical segment. [SEE LETTER P.83](#)

ARMIN P. MOCZEK

Understanding the origin of complex traits is among the most enduring puzzles in evolutionary biology. On the one hand, evolution operates within a framework of descent with modification — everything new must come from something old. On the other hand, structures such as the eye, the wing and the turtle's shell stand out because they lack obvious correspondence to the old. On page 83 of this issue, Prud'homme *et al.*¹ address this puzzle by connecting a complex and highly diverse trait — the helmet of membracid treehoppers — to its origins in both development and evolution.

Treehoppers are insects that would resemble miniature cicadas were it not for the presence of the helmet (Fig. 1). This structure appears to reside on top of the animal's thorax, and extends dorsally, and in remarkably varied ways, to mimic thorns, animal droppings or aggressive ants. Entomologists joke that some treehoppers use their helmets to send signals to their home planet, so other-worldly is their appearance.

Helmets have been interpreted as an extension of the pronotum, the dorsal portion of the first segment of the three-segmented thorax shared by all insects². The thorax is a defining feature of insects, bearing a pair of legs on each of its three segments and, in most insect orders, a pair of wings on the second and third segments (but not on the first, the prothorax). We have long known from fossil evidence that insects arrived at this organization following a period of progressive loss of wings or wing-like appendages from all abdominal segments, as well as from the first thoracic segment³ (Fig. 2). More recently, developmental studies have shown that this loss has been achieved through the evolution of inhibitory mechanisms that prevent the formation of wings in inappropriate segments. For instance, one of the many functions of a gene called *Sex combs reduced* (*Scr*) is to mediate the inhibition of wing formation in the first thoracic segment of insects⁴, including the order Hemiptera, to which treehoppers belong⁵.

Enter the treehopper *Publilia modesta* and its helmet. Through careful analysis of this structure's anatomy, placement and

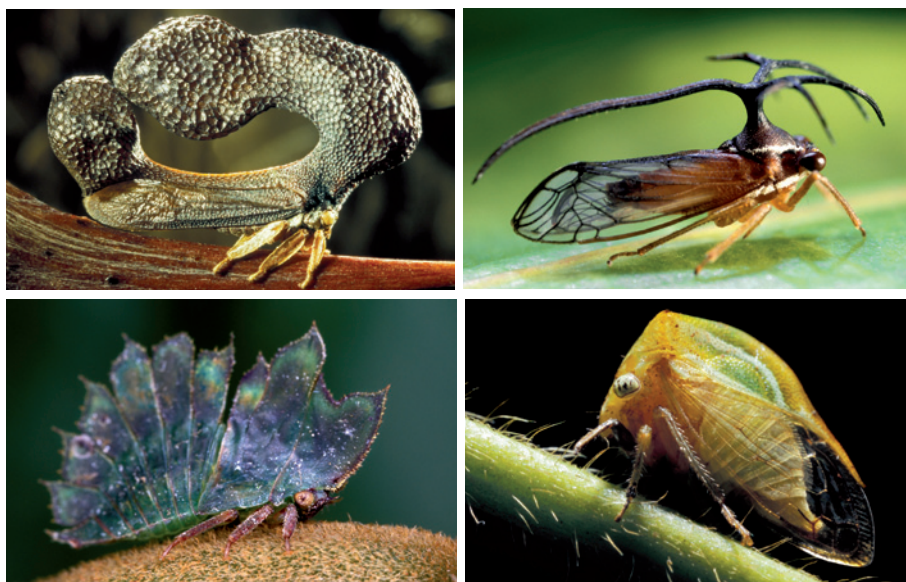


Figure 1 | The exuberance of treehopper helmets. Clockwise from top left: *Cladonota benitezi*; *Umbelligerus peruviansis*; *Nassunia binotata*; and a nymph of a *Cymbomorpha* species. Helmets are generally thought to aid in camouflage by disrupting the animal's shape and outline, or by mimicking thorns, animal droppings or aggressive ants and wasps. Further examples are shown on the cover of this issue, and in Figure 1 of Prud'homme and colleagues' paper¹.

attachment to the thorax, Prud'homme *et al.*¹ discovered that the helmet may not be a mere extension of the pronotum. Instead, it is attached bilaterally to the thorax by paired articulations reminiscent of joints, much like regular wings. Moreover, when they examined its early developmental stages, the authors found that the helmet forms from paired buds — again, much like wings. The expanding buds subsequently fuse along the midline, creating the continuous helmet. Study of the expression of one gene, *nubbin*, normally specific to insect wing development, and two genes specific to appendage formation in general, provided additional evidence that helmet development may rely on developmental mechanisms involved in the formation of wings.

Combined, these observations suggested that treehoppers evolved a way to develop a wing-like structure using a developmental program shared by traditional wings, but in a place in which wing development is typically inhibited in modern winged insects. Prud'homme and colleagues' investigation

of *Scr* revealed that the gene is still expressed in the prothorax of treehoppers and is able to repress wing formation when transformed into *Scr*-deficient fruit flies. This implies that wing development in the first thoracic segment of treehoppers was not made possible simply by the loss of the inhibitory ability of *Scr*, but through some unknown mechanisms operating downstream.

The study by Prud'homme *et al.*¹ is noteworthy for several reasons. First, it illustrates how, to this day, careful developmental observations can set the stage for startling discoveries. Generations of entomologists have studied treehopper diversity, but research into development has a way of revealing evolution hidden from the study of adults. Second, as with so many studies, it raises as many questions as it answers. Although the morphological observations provide strong evidence that the helmet is a modified wing, the developmental genetic data are modest and correlational: expression patterns can suggest, but not prove, function. And the mechanisms that permit wing-like development in the

P. LANDMANN/SPL



Figure 2 | A wing-bearing first thoracic segment. As shown in this line drawing of a fossil of an extinct species (*Stenodyctya lobata*), expression of the wing-development program in the first thoracic segment (arrow) was common in early insects. In extant winged insects, wings are borne only on the second and third thoracic segments, with wing development on the first segment being suppressed. Prud'homme *et al.*¹ provide evidence that treehoppers have overcome such suppression to produce their helmets. (Drawing reproduced from Fig. 6.17 of ref. 3.)

BIOCHEMISTRY

Life imitates art

The biosynthetic route to a naturally occurring insecticide, spinosyn A, has been established. One of the enzymes involved might catalyse a reaction that, although widely used by chemists, has proved elusive in nature. SEE LETTER P.109

WENDY L. KELLY

The Diels–Alder reaction is a powerful instrument in the synthetic organic chemist's toolkit¹. A variant of '[4+2] cycloaddition' reactions, the Diels–Alder reaction forges two carbon–carbon single bonds in the process of making a cyclohexene ring — a six-membered carbon ring possessing a carbon–carbon double bond. A biochemical

equivalent of this process has been invoked as a crucial step in the biosynthesis of many naturally occurring molecules, but the roster of enzymes that clearly catalyse transformations consistent with the Diels–Alder reaction has been limited. What's more, the enzymes on that list mediate sequences of reactions, of which the putative Diels–Alder reaction is just one, thereby confusing efforts to study biological cycloadditions.

On page 109 of this issue, Kim *et al.*² identify an enzyme whose sole function is to catalyse the formation of a cyclohexene, a process consistent with a Diels–Alder reaction. This transformation, along with the others detailed in the authors' report, is a critical step in the biosynthesis of spinosyn A, a commercially useful and environmentally friendly insecticide.

Spinosyn A belongs to the polyketide family of natural products, and is produced by fermentation of the bacterium *Saccharopolyspora spinosa*³. The molecular backbone of spinosyn A is a complex framework: a large 'lactone' ring is fused to a highly unusual system called a perhydro-*as*-indacene, which consists of three smaller rings (Fig. 1a). During the biosynthesis of spinosyn A, a polyketide synthase enzyme assembles the molecule's carbon backbone, initially generating a single large ring (a macrocycle). Later in the synthesis, the macrocycle is converted into the multi-ring system and glycosyltransferase enzymes attach carbohydrate groups to the macrocyclic scaffold.

Although a [4+2] cycloaddition has been proposed as a key step in the installation of spinosyn A's fused-ring system, the exact point at which this occurs was uncertain. It was suggested that the system is generated when a series of carbon–carbon bonds form as bridges across a macrocyclic intermediate consisting of only one ring. This proposal was strengthened by the discovery that SpnJ — an enzyme involved in spinosyn A biosynthesis — uses an unbridged macrocyclic precursor of spinosyn A as its substrate^{4,5}. Bridge-forming reactions must therefore occur after the macrocycle has been formed. The bridge-forming process could follow at least two paths, which would differ according to whether the proposed [4+2] cycloaddition precedes or follows formation of the bridge between positions 3 and 14 of spinosyn A (Fig. 1a shows how the atoms in spinosyn A are numbered).

Kim *et al.*² now reveal the full sequence of reactions that proceed from an unbridged macrocyclic intermediate to the characteristic fused-ring system of spinosyn A. They find that a [4+2] cycloaddition reaction, catalysed

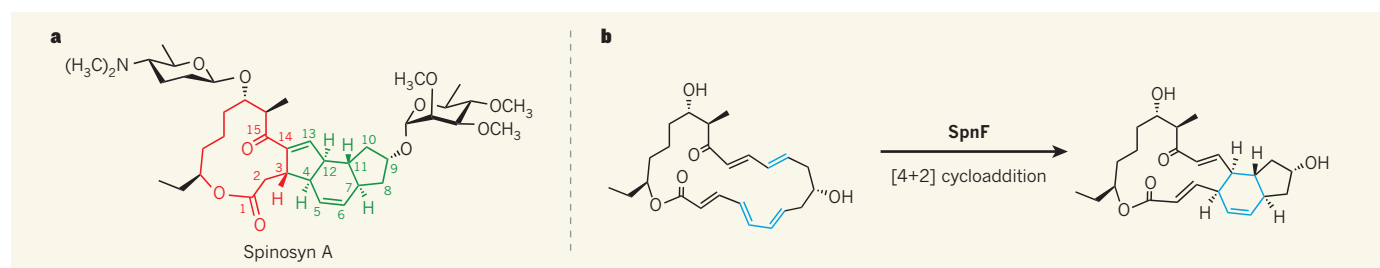


Figure 1 | The biosynthesis of spinosyn A. a, Kim *et al.*² have worked out the biosynthetic pathway for spinosyn A, a naturally occurring insecticide. The core structure contains a macrocyclic lactone (red) fused to a perhydro-*as*-indacene system (green). Part of the numbering system used to identify the

atoms in the molecule is shown. b, A [4+2] cycloaddition reaction catalysed by the enzyme SpnF is a key step in the formation of the perhydro-*as*-indacene. The reacting parts of the starting material, and the cyclohexene ring formed in the product, are highlighted in blue.

by the enzyme SpnF, does indeed occur. More specifically, the reaction constructs two of the bridges in the spinosyn framework — one between positions 7 and 11, and the other between positions 4 and 12 — as a first step in the synthesis of the perhydro-*as*-indacene system (Fig. 1b). A glycosyltransferase, SpnG, next appends a carbohydrate to the resulting scaffold, before the final bridge between carbon atoms at positions 3 and 14 is introduced by the enzyme SpnL. Intriguingly, this biochemical route to the perhydro-*as*-indacene framework of spinosyn A is the same as that used by synthetic chemists William Roush and co-workers in their laboratory preparation of the molecule⁶ — an example of life imitating the art of synthetic chemistry.

Before Kim and colleagues' report², only four enzymes had been identified that seemed to mediate a [4+2] cycloaddition to form a cyclohexene ring: lovastatin nonaketide synthase⁷ (LovB), solanapyrone synthase⁸, riboflavin synthase⁹ and macrophomate synthase¹⁰. Each of these enzymes catalyses at least one other chemical transformation in addition to a cycloaddition. Perhaps LovB is the most multi-functional — as well as mediating a cycloaddition, it harbours seven functional domains, each attributed to separate chemical processes, and many of which are used iteratively to assemble a polyketide backbone containing 18 carbon atoms^{7,11}.

Because SpnF effects only a [4+2] cycloaddition, a detailed examination of its reaction mechanism will be uncomplicated by other transformations. So far, Kim *et al.*² have established that SpnF is a genuinely catalytic protein that enhances the rate of the non-enzymatic cycloaddition reaction 500-fold. But a fundamental remaining question is whether the SpnF-mediated cycloaddition is a true Diels–Alder reaction.

The hallmark of Diels–Alder [4+2] cycloadditions is that they are concerted — they proceed without forming any transient intermediates en route to the final product. Despite the identification of five candidate enzymes for [4+2] cycloadditions, none of these has yet been proved to mediate a concerted reaction mechanism. If the reaction catalysed by SpnF does turn out to be concerted, it would be the first example of a 'Diels–Alderase', an enzyme that catalyses a Diels–Alder reaction. The challenge now for Kim *et al.* is to perform a detailed mechanistic study of the SpnF-catalysed reaction to determine whether or not it proceeds through intermediates.

Of the five potential Diels–Alderases identified to date, macrophomate synthase has been subjected to the most detailed mechanistic analysis. In this case, there is mounting evidence^{12,13} that, rather than serving as a bona fide Diels–Alderase, the enzyme directs a two-step [4+2] cycloaddition.

Nevertheless, it is enticing to speculate that SpnF, and the other three enzymes, catalyse concerted reactions. The fact that SpnF catalyses only a [4+2] cycloaddition greatly simplifies the analyses required to address the Diels–Alderase question. As a result, SpnF may eventually offer the clearest answer. ■

Wendy L. Kelly is in the School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332-0400, USA. e-mail: wendy.kelly@chemistry.gatech.edu

- Huisgen, R. *Angew. Chem. Int. Edn* **7**, 321–328 (1968).
- Kim, H. J., Ruszczycky, M. W., Choi, S., Liu, Y. & Liu, H. *Nature* **473**, 109–112 (2011).
- Kirst, H. A. *J. Antibiot.* **63**, 101–111 (2010).
- Waldron, C. *et al. Chem. Biol.* **8**, 487–499 (2001).
- Kim, H. J., Pongdee, R., Wu, Q., Hong, L. & Liu, H. *J. Am. Chem. Soc.* **129**, 14582–14584 (2007).
- Mergott, D. J., Frank, S. A. & Roush, W. R. *Proc. Natl Acad. Sci. USA* **101**, 11955–11959 (2004).
- Auclair, K. *et al. J. Am. Chem. Soc.* **122**, 11519–11520 (2000).
- Kasahara, K. *et al. ChemBioChem* **11**, 1245–1252 (2010).
- Kim, R.-R. *et al. J. Am. Chem. Soc.* **132**, 2983–2990 (2010).
- Watanabe, K., Mie, T., Ichihara, A., Oikawa, H. & Honma, M. *J. Biol. Chem.* **275**, 38393–38401 (2000).
- Ma, S. M. *et al. Science* **326**, 589–592 (2009).
- Guimarães, C. R. W., Udier-Blagović, M. & Jorgensen, W. L. *J. Am. Chem. Soc.* **127**, 3577–3588 (2005).
- Serafimov, J. M., Gillingham, D., Kuster, S. & Hilvert, D. *J. Am. Chem. Soc.* **130**, 7798–7799 (2008).

enhance contraction at a given ventricular volume — were sought in order to enhance the Ca²⁺ signal that activates contraction. But many of the early drugs (for example, digitalis) could overload cardiac muscle cells with Ca²⁺, increasing both energy consumption and the risk of arrhythmias. Indeed, several such agents — including phosphodiesterase inhibitors — mimicked the effect of sympathetic stimulation via β -adrenergic receptors, which greatly increases energy consumption by the heart during the normal physiological 'fight-or-flight' response. In heart failure, this increase in energy consumption can worsen a patient's prognosis.

Most of the current standard-of-care drugs used for patients with chronic heart failure, including β -adrenergic-receptor blockers (β -blockers), angiotensin-converting-enzyme inhibitors (ACE inhibitors) and angiotensin II-receptor blockers (ARBs), are not inotropic drugs; instead, they block neurohumoral signalling by adrenergic and renin–angiotensin pathways. Heart failure is accompanied by a neurohumoral storm that activates these pathways, probably as an initially adaptive response that turns maladaptive by fuelling progressive remodelling and dysfunction. Blocking these pathways can partially break this vicious circle and slow the progress of heart failure¹.

Newer Ca²⁺-related inotropic strategies are

TRANSLATIONAL MEDICINE

To the rescue of the failing heart

Heart failure is characterized by weakened contractions of heart muscle. A drug that directly activates the key force-generating molecule in this muscle may be a valuable tool to strengthen the failing heart.

DONALD M. BERS & SAMANTHA P. HARRIS

Heart failure affects tens of millions of people worldwide, with patients' prognosis often being a bleak five-year survival from the time of diagnosis¹. Patients die because of a vicious circle of progressive weakening of the heart leading to cardiac remodelling, which further weakens it and can also cause deadly arrhythmias. If the failing heart could be strengthened, the outcome might be more favourable. Writing in *Science*, Malik *et al.*² describe a small-molecule drug — omecamtiv mecarbil — that selectively enhances the activity of the motor protein myosin, the main force-generating protein of the heart.

At each heartbeat, a specialized intracellular

organelle, the sarcoplasmic reticulum, releases calcium ions (Ca²⁺) into the cytoplasm of the heart-muscle cells in a synchronized manner (Fig. 1). The Ca²⁺ activates myofilaments — organized structures in the cytoplasm composed of interdigitating filaments of either actin or myosin proteins. On activation, each myosin filament simultaneously grabs and pulls on an actin filament, in a process that uses the cellular energy molecule ATP. The coordinated contractile activity of the myofilaments develops the forceful muscle contraction that ejects blood from the heart³. In heart failure, a reduced amount of Ca²⁺ is available for release by the sarcoplasmic reticulum, contributing to weaker myofilament activation and contraction.

Historically, inotropic drugs — drugs that

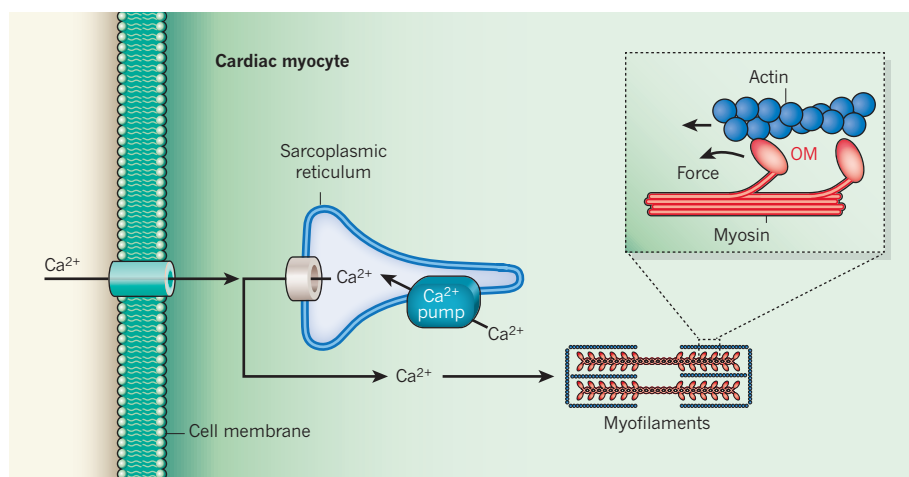


Figure 1 | How omecamtiv mecarbil functions. In heart muscle cells, Ca^{2+} influx from outside the cell triggers Ca^{2+} release from the sarcoplasmic reticulum. This in turn activates contraction by enabling myosin molecules to grab and pull on the actin filament. Reuptake of Ca^{2+} by the Ca^{2+} pump allows heart relaxation between beats. Increasing either the amount of Ca^{2+} released or the myofilaments' response to Ca^{2+} — as induced by the drug omecamtiv mecarbil (OM)² — can enhance contractility.

more precisely focused on molecular targets with the aim of enhancing sarcoplasmic-reticulum function while minimizing the energetic disadvantage and arrhythmia risk of the older drugs. For instance, gene therapy aims to increase expression in the sarcoplasmic reticulum of the Ca^{2+} pumps, which are downregulated in heart failure⁴. Other examples are drugs that either block pathological Ca^{2+} leak from the sarcoplasmic reticulum or stimulate Ca^{2+} uptake by this organelle⁵. These drugs might more selectively boost the transient increase in cytoplasmic Ca^{2+} levels without causing arrhythmia and with limited energetic consequences. So there is also hope for refinement of this strategy.

Malik *et al.*² find that omecamtiv mecarbil — also an inotropic drug — increases heartbeat strength by selectively enhancing the ability of the myosin molecule to generate force (Fig. 1). However, rather than boosting Ca^{2+} release, it jumps downstream and allows generation of greater force for the same Ca^{2+} signal. Targeting the final step of force production is a big advantage of this approach, because it potentially avoids unintended side effects typical of other upstream modulators of Ca^{2+} handling or neurohumoral signalling.

Indeed, the authors report that omecamtiv mecarbil enhances cardiac output without appreciably altering consumption of oxygen and ATP by the heart. This is presumably because any extra ATP is used right at the force-generating step, rather than being also used to transport Ca^{2+} into the sarcoplasmic reticulum or out of the cell, or via altered metabolism. As the heart weakens, it receives less nutritive, oxygen-rich blood (that is, the heart pumps blood through its own coronary arteries), which further limits cardiac contraction. By augmenting force while avoiding extra energetic costs, omecamtiv mecarbil increases the apparent efficiency of cardiac contraction and

preserves the energy supply–demand balance.

Omecamtiv mecarbil belongs to the class of drug that enhances contractile protein responses to Ca^{2+} — such as levosimendan⁶ — by increasing the force produced for a given level of Ca^{2+} release. But two aspects of Malik and co-workers' study are particularly noteworthy.

First, cardiac myosin is a new drug target and, although quite promising, the drug might have unintended side effects. Between beats, the heart must relax completely (the diastolic phase) to allow refilling with blood and to provide adequate oxygen flow to the heart muscle (most coronary blood flow is between beats). Because omecamtiv mecarbil prolongs contraction time, diastolic filling may be compromised, especially at higher heart rates. Drugs that allow significant force generation by myofilaments at diastolic Ca^{2+} levels can also impede ventricular refilling, and elevate cardiac stiffness and diastolic energy consumption. This potential limitation of omecamtiv mecarbil should be further assessed.

Second, although this drug is highly specific for cardiac myosin, slow skeletal-muscle fibres also use the same myosin isoform as in cardiac muscle. Consequently, omecamtiv mecarbil may cause stronger, more sustained contractions in slow-twitch muscles too. If so, it is tempting to speculate that the drug could find additional therapeutic or performance-related applications — for instance, in strengthening diaphragm muscles of patients on ventilators. The prospect of developing other small-molecule activators that specifically target fast skeletal myosins could hold similar promise for augmenting force in the muscle wasting that occurs in cancer or ageing.

Agents such as omecamtiv mecarbil could certainly contribute to future therapy for those who have heart failure. Complex and pervasive as heart failure is, so, fortunately, is the range of



50 Years Ago

P. M. Borisov has outlined a project, in the ... *Literaturnaya Gazeta*, of a 90 km. long dam across the Bering Strait equipped with powerful pumps pumping cold Arctic Ocean water into the Pacific Ocean at the rate of 500 km.³ in 24 hours. Such a project ... would increase the flow of warm Atlantic Ocean water into the Arctic Ocean and change the climate of the Arctic regions. This project is criticized by D. A. Drogaitzev... [who] argues that such a project would displace the locus of Atlantic Ocean cyclones to the region of the Barentz Sea. Such a displacement would certainly change the climate of Northern Europe and Western Siberia, but this change will produce colder winters and hotter summers and will lead to the displacement of the desert belt from the region of North Africa and Central Asia to the north of Europe.
From *Nature* 6 May 1961

100 Years Ago

Three letters have recently appeared in *The Times* ... relating to a mysterious heraldic animal known as the “jall” or “eall”, of which the effigy has been recognised in St. George's Chapel, Westminster ... Although described as having horns, tusks, and a short fluffy tail, the jall has been identified with the goat ... In an old document ... the eall is stated to be as large as a horse, with a tail like that of an elephant, goat-like jaws, and horns capable of movement, its colour being black. Other accounts state, however, that it has jaws like a wild boar and cloven hoofs. It may be suggested, if the beast ever had corporeal existence, that the African wart-hog may have formed the original type, that animal having a black hide, cloven hoofs, an elephant-like tail, large tusks, and big face-warts which might perhaps be regarded as elastic horns.
From *Nature* 4 May 1911

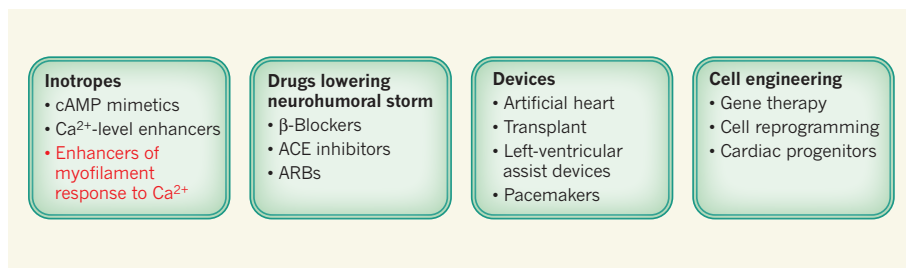


Figure 2 | Treating heart failure. There is an extensive array of therapeutic strategies for heart failure. Omecamtiv mecarbil, the subject of Malik and colleagues' investigation², is an inotropic drug (red). ARBs, angiotensin II-receptor blockers.

therapeutics being developed and used against it¹ (Fig. 2). Although cardiac transplantation is the only real cure for heart failure, artificial hearts and left-ventricular assist devices are beneficial, at least as bridges to transplantation and perhaps even as 'destination therapy'. Implantable devices such as pacemakers and resynchronization devices are also useful for treating heart failure, but devices

and surgical interventions can be costly.

There is also great promise in emerging genetically based therapeutics that aim to replace or reprogram cardiac myocytes in order to boost heart function. For example, selective gene therapy targeted to cardiac myocytes might be able to break the neurohumoral storm and enhance myocyte contraction with fewer whole-body side effects than other therapies.

Reprogramming of stem cells (either embryonic or inducible pluripotent), together with recruitment of cardiac progenitor cells to become functionally integrated muscle cells that can replace heart muscle lost to infarction, are promising areas under intensive study. The old notion that one cannot grow new heart cells in adulthood is probably incorrect. ■

Donald M. Bers is in the Department of Pharmacology and **Samantha P. Harris** is in the Department of Neurobiology, Physiology and Behavior, University of California, Davis, Davis, California 95616, USA.
e-mail: dmbars@ucdavis.edu

1. Tang, W. H. & Francis, G. S. *J. Am. Coll. Cardiol.* **55**, 688–696 (2010).
2. Malik, F. I. *et al. Science* **331**, 1439–1443 (2011).
3. Bers, D. M. *Nature* **415**, 198–205 (2002).
4. Kawase, Y., Ladage, D. & Hajjar, R. J. *J. Am. Coll. Cardiol.* **57**, 1169–1180 (2011).
5. Kushnir, A. & Marks, A. R. *Adv. Pharmacol.* **59**, 1–30 (2010).
6. Kass, D. A. & Solaro, R. J. *Circulation* **113**, 305–315 (2006).

METROLOGY

Filtering noise with a quantum probe

In the science of measurement, increasing the sensitivity to the quantity being measured while minimizing the susceptibility to noise is a challenge. A technique demonstrated with a single electron spin may help to tackle it. SEE LETTER P.61

JOHN J. BOLLINGER

Applications in both fundamental and applied science require ever greater sensitivity and higher spatial resolution for measurements of physical quantities such as magnetic and electric fields. The most sensitive and smallest measurement probes are inherently quantum mechanical. Examples include superconducting quantum-interference devices (SQUIDs)¹ and devices based on a few electron spins — or even a single spin². However, greater susceptibility to noise usually accompanies extreme sensitivity, and so one of the challenges for metrologists is to separate a weak signal from large background noise.

On page 61 of this issue, Kotler and colleagues³ describe a general technique in which a quantum probe is used to separate noise from the signal being measured, and they demonstrate it experimentally using a probe consisting of the spin of a valence electron of an individual atomic particle (a single strontium ion). The technique requires a controlled modulation of the quantity to be measured and a corresponding controlled manipulation of

the quantum probe. It is reminiscent of noise-filtering techniques developed decades ago for classical signals and probes, but is described here for a general quantum probe for the first time.

More than half a century ago, Robert Dicke invented the lock-in amplifier⁴. This powerful tool is now used extensively in all branches of experimental science to extract signal from a noisy background. As an example of lock-in detection, consider the measurement of a weak fluorescent signal in the presence of strong background light. If the clever experimentalist can devise a way to periodically modulate the weak fluorescent signal at a frequency f_m , for example by modulating the number of fluorescing molecules, then the overall detected signal will contain a contribution whose time dependence is given by a known sinusoidal reference signal of frequency f_m . The lock-in amplifier electronically multiplies the overall signal (for example, the voltage output of the sensor with which the fluorescence is detected) and the reference signal, and averages the result for a period of time. An output is therefore generated that is proportional to the signal components around

some narrow band of frequencies centred at f_m . Noise tends to be spread across a broad range of frequencies, and the lock-in amplifier filters out noise at frequencies other than f_m . By choosing f_m judiciously, the signal-to-noise ratio of the measurement can be significantly improved.

The classical lock-in amplifier is based on the nonlinear process of multiplying the output of the sensor and the reference signal. However, quantum dynamics is described by a linear differential equation (the Schrödinger equation), and so it is not immediately clear how the concept of lock-in detection could be generalized to a quantum-mechanical probe. Kotler and colleagues³ show that the application of operations that do not commute with the quantum-mechanical operators describing the detected signal and noise, along with a synchronous modulation of the signal to be measured, provides a form of quantum lock-in detection.

This abstract idea is actually familiar to anyone acquainted with the concept of spin echoes, a ubiquitous technique in nuclear magnetic resonance⁵. Consider a single electron or nuclear spin that is set precessing about an externally applied magnetic field. As a result of magnetic-field fluctuations, the spin accumulates some unknown precession. For slow noise fluctuations, this unknown precession can be reversed by means of a spin-echo pulse — a quick 180° rotation about an axis orthogonal to the magnetic field. Mathematically, rotations about orthogonal axes do not commute. Spin echo is a simple example of a more general class of technique called dynamical decoupling⁶, which relies on stringing together many spin-echo pulses in succession.

Dynamical-decoupling sequences improve the coherence of quantum systems by acting

as high-pass filters, removing the effects of environmental fluctuations (noise) across a wide spectral bandwidth. However, such suppression of noise comes at a price in metrology experiments, because the intrinsic high-pass filtering prevents certain quantities from being measured. For example, because the spin echo 'erases' the accumulation of unwanted precession in a quantum system, one cannot measure the rate at which any precession accumulates.

To get around this, Kotler *et al.*³ exploit a quirk of dynamical-decoupling sequences: high-frequency noise is not passed uniformly, allowing the authors to home in on the quantum effects of a desired signal in a narrow-frequency passband. The passband is controlled by the dynamical-decoupling sequence of the spin-echo pulses they apply. Changing the periodicity of the applied pulses tunes the central frequency of this band, and by synchronously modulating the signal of interest, the

quantum lock-in amplifier preserves the signal while the dynamical decoupling filters the noise — the authors can have their cake and eat it too.

They experimentally demonstrate the quantum lock-in technique using a probe consisting of the valence-electron spin of a singly ionized strontium atom that is laser-cooled and stored in an electromagnetic trap. The spin-flip frequency of the unpaired valence electron is sensitive to an applied magnetic field, and the authors demonstrate a magnetic-field sensitivity of 15 picoteslas of magnetic-field strength in a 1-second measurement period — a record for a single-spin probe. In addition, they apply the technique to measure small shifts in the spin-flip frequency of the valence electron caused by a weak applied laser field. This demonstration is particularly intriguing because it provides a way to use dynamical decoupling to stabilize the

frequency of a laser to that of an atomic transition. Lasers stabilized to narrow-linewidth atomic transitions currently provide the world's most stable atomic clocks^{7,8}. ■

John J. Bollinger is in the Time and Frequency Division, National Institute of Standards and Technology, Boulder, Colorado 80305, USA.

e-mail: john.bollinger@boulder.nist.gov

1. Bending, S. J. *Adv. Phys.* **48**, 449–535 (1999).
2. Taylor, J. M. *et al. Nature Phys.* **4**, 810–816 (2008).
3. Kotler, S., Akerman, N., Glickman, Y., Keselman, A. & Ozeri, R. *Nature* **473**, 61–65 (2011).
4. Horowitz, P. & Hill, W. *The Art of Electronics* 1031 (Cambridge Univ. Press, 1989).
5. Haeberlen, U. *High Resolution NMR in Solids* (Adv. Magn. Res. Ser.) (Academic, 1976).
6. Biercuk, M. J. *et al. Nature* **458**, 996–1000 (2009).
7. Chou, C. W., Hume, D. B., Koelemeij, J. C. J., Wineland, D. J. & Rosenband, T. *Phys. Rev. Lett.* **104**, 070802 (2010).
8. Ludlow, A. D. *et al. Science* **319**, 1805–1808 (2008).

Terashima *et al.*¹ now suggest an original solution to this problem. Instead of synthesizing a polymer that has a complex, three-dimensional topology, they investigated whether a single linear polymer chain can be folded to make an enzyme-like object for catalysis. To do this, the authors carefully designed a polymer chain that was constructed from three different monomers (Fig. 1a): a hydrophilic monomer that contained a water-soluble group; another that bore a self-assembling motif; and a third monomer that contained a diphenylphosphine ligand, which forms a catalytic complex with ruthenium ions.

These monomers were not, however, randomly incorporated into a polymer backbone. Using an approach known as living radical polymerization⁸, the authors controlled the locations of the different monomers in the polymer chains⁹. For instance, they specifically incorporated the ruthenium-binding monomers into the middle of the chains, whereas the other types of monomer were distributed along the whole length of the chains. What's more, because the polymerization reaction required a ruthenium catalyst, the diphenylphosphine groups in the chains formed complexes with ruthenium ions from that catalyst. The arrangement of monomers in the resulting chains caused the molecules to fold up in water (Fig. 1b), as a result of intramolecular hydrophobic and hydrogen-bonding interactions. In particular, the self-assembling units incorporated into the polymer formed compact helical structures, so that the linear macromolecules collapsed like supramolecular accordions.

Terashima *et al.* found that, as hoped, their macromolecules folded into unimolecular objects in which a catalytically active inner region (the domain containing ruthenium complexes) was stabilized by a hydrophilic shell. This compartmentalization was thus a good — albeit simplified — mimic of the

MATERIALS CHEMISTRY

Catalytic accordions

Single chains of a specially designed polymer fold up in water to form an encapsulated catalytic chamber. This supramolecular assembly strategy mimics the one used by enzymes in nature.

NICOLAS GIUSEPPONE
& JEAN-FRANÇOIS LUTZ

The catalytic properties of an enzyme result from the three-dimensional folding of a single protein chain, which brings together a well-defined set of amino-acid residues to form the enzyme's active site. This pocket is a highly organized domain that binds tightly and selectively to the enzyme's substrate, which becomes trapped and polarized in a network of supramolecular interactions. In this way, active sites lower the energy of transition states for reactions, so that products form up to billions of times faster than in the uncatalysed reactions. A challenge for chemists has been to devise systems that mimic enzyme activity, and a breakthrough has now been reported by Terashima *et al.*¹ in the *Journal of the American Chemical Society*. They have synthesized a polymer, single chains of which fold in water to form an inner compartment that acts, through its supramolecular structure, as an 'active site' for a catalytic reaction.

Supramolecular chemistry is fundamental to catalysis, because the transition states of chemical reactions represent a special class of supramolecular complex in which some covalent bonds are being formed while others are being broken. What's more, numerous self-assembled supramolecular objects have been

designed to act as catalysts², in particular by acting as templates that bring reagents together to react. Examples of these include cages or capsules made of discrete small molecules³ or proteins⁴, and multi-component matrices⁵ such as micelles or vesicles, made of surfactants or polymers. But these self-assemblies are relatively poor catalysts in comparison with highly organized enzymes.

Other options for developing artificial enzymes have therefore been studied. For instance, it is possible to prepare fully synthetic enzymes from amino acids by using well-established chemistry to make polypeptide fragments, and then joining the fragments together to construct proteins in so-called ligation reactions⁶. However, such approaches are still rather challenging and time-consuming.

Simpler alternatives are obviously required. Given that enzymes are macromolecular, the idea of performing catalytic reactions in other discrete macromolecular entities, such as polymer molecules, seems logical. Macromolecular objects made from branched polymers have received much attention in this regard⁷, because they contain isolated domains that could be used as catalytic active sites. But the three-dimensional structures of branched polymers are not obtained through straightforward supramolecular folding, as is the case for enzymes. They are instead the topological result of complex synthetic routes.

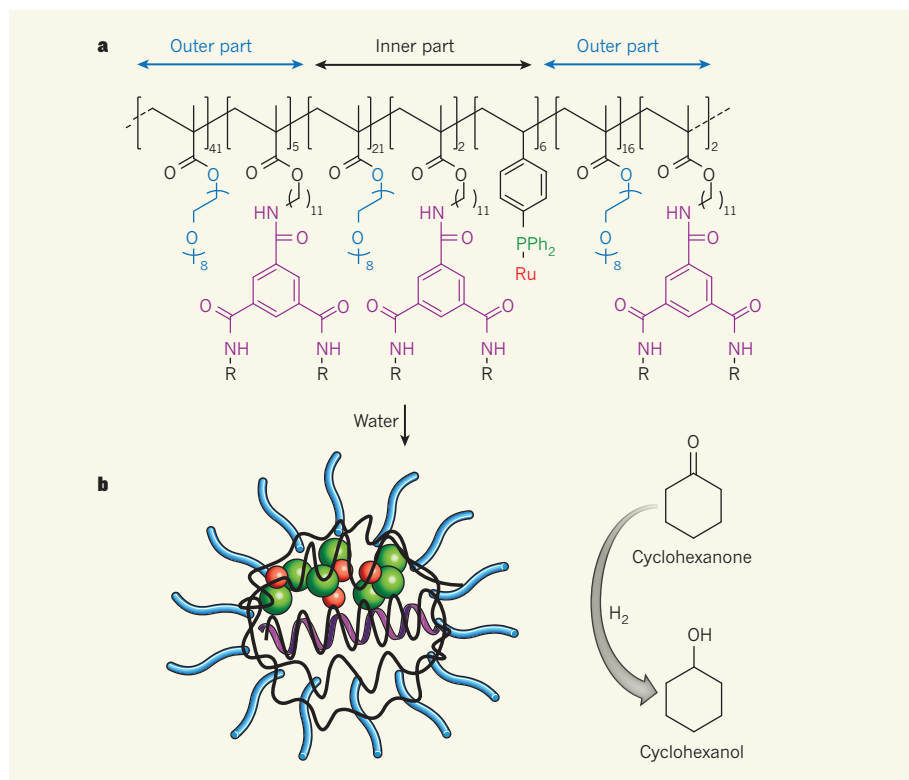


Figure 1 | Enzyme-like polymer folding. **a**, Terashima *et al.*¹ have made a polymer in which hydrophilic groups (blue), self-assembling groups (purple) and ligand groups (green) are attached to the polymer's hydrophobic 'backbone' (black). The ligands form complexes with ruthenium ions (red). R is a hydrocarbon group. **b**, In water, intramolecular interactions cause single chains of the polymer to adopt an enzyme-like structure: the self-assembling groups form helices, and the ruthenium ions bound to ligands become surrounded by a hydrophobic shell (black), which in turn is surrounded by hydrophilic groups (blue). These folded structures act as ruthenium catalysts for the reaction in which cyclohexanone reacts with hydrogen to form cyclohexanol.

structural organization found in biological enzymes. The authors went on to demonstrate that their polymeric objects catalyse reactions in which hydrogen atoms are added to ketone molecules (Fig. 1b), with a catalytic activity comparable to that of previously reported water-soluble ruthenium complexes.

The folding of Terashima and colleagues' polymers is not as defined as that in naturally occurring enzymes, but it is certainly more ordered than the simple random coiling found in traditional polymers. It is likely that greater structural definition in polymers will be possible in the future, using methods that improve the precision of supramolecular folding^{8,9}. This would enable artificial enzymes to be made that have higher catalytic activity and substrate selectivity than the authors' current system, and might even allow allosteric recognition — modulation of the artificial enzyme by ligand binding at sites other than the active site.

Terashima and colleagues' work¹ is at the forefront of polymer science. Their findings clearly demonstrate the potential of single-chain polymer folding in catalysis, and illustrate how macromolecular and supramolecular chemistry can converge in the fabrication of functional molecular assemblies in general, something that might be

useful in other applications. For example, the folding of synthetic linear polymer chains is a robust, but as-yet under-explored, option for materials design. Their work might also help in the development of methods for predicting how polymers fold — an unsolved problem that is especially important for proteins, the keystones that link the genetic code to its expressed biological functions. ■

Nicolas Giuseppone is in the Faculty of Chemistry, University of Strasbourg, and at the Institut Charles Sadron, UPR22-CNRS, 67034 Strasbourg, France. **Jean-François Lutz** is at the Institut Charles Sadron.
e-mails: giuseppone@unistra.fr;
jflutz@unistra.fr

1. Terashima, T. *et al.* *J. Am. Chem. Soc.* **133**, 4742–4745 (2011).
2. Breslow, R. *Science* **218**, 532–537 (1982).
3. Kang, J. & Rebek, J. *Nature* **385**, 50–52 (1997).
4. Douglas, T. & Young, M. *Nature* **393**, 152–155 (1998).
5. Dwaars, T., Paetzold, E. & Oehme, G. *Angew. Chem. Int. Edn* **44**, 7174–7199 (2005).
6. Kent, S. B. H. *Chem. Soc. Rev.* **38**, 338–351 (2009).
7. Hecht, S. & Fréchet, J. M. J. *Angew. Chem. Int. Edn* **40**, 74–91 (2001).
8. Ouchi, M., Terashima, T. & Sawamoto, M. *Chem. Rev.* **109**, 4963–5050 (2009).
9. Badi, N. & Lutz, J.-F. *Chem. Soc. Rev.* **38**, 3383–3390 (2009).

Mapping and analysis of chromatin state dynamics in nine human cell types

Jason Ernst^{1,2}, Pouya Kheradpour^{1,2}, Tarjei S. Mikkelsen¹, Noam Shores¹, Lucas D. Ward^{1,2}, Charles B. Epstein¹, Xiaolan Zhang¹, Li Wang¹, Robbyn Issner¹, Michael Coyne¹, Manching Ku^{1,3,4}, Timothy Durham¹, Manolis Kellis^{1,2*} & Bradley E. Bernstein^{1,3,4*}

Chromatin profiling has emerged as a powerful means of genome annotation and detection of regulatory activity. The approach is especially well suited to the characterization of non-coding portions of the genome, which critically contribute to cellular phenotypes yet remain largely uncharted. Here we map nine chromatin marks across nine cell types to systematically characterize regulatory elements, their cell-type specificities and their functional interactions. Focusing on cell-type-specific patterns of promoters and enhancers, we define multicell activity profiles for chromatin state, gene expression, regulatory motif enrichment and regulator expression. We use correlations between these profiles to link enhancers to putative target genes, and predict the cell-type-specific activators and repressors that modulate them. The resulting annotations and regulatory predictions have implications for the interpretation of genome-wide association studies. Top-scoring disease single nucleotide polymorphisms are frequently positioned within enhancer elements specifically active in relevant cell types, and in some cases affect a motif instance for a predicted regulator, thus suggesting a mechanism for the association. Our study presents a general framework for deciphering *cis*-regulatory connections and their roles in disease.

A major challenge in biology is understanding how a single genome can give rise to an organism comprising hundreds of distinct cell types. Much emphasis has been placed on the application of high-throughput tools to study interacting cellular components¹. The field of systems biology has exploited dynamic gene expression patterns to reveal functional modules, pathways and networks². Yet *cis*-regulatory elements, which may be equally dynamic, remain largely uncharted across cellular conditions.

Chromatin profiling provides a systematic means of detecting *cis*-regulatory elements, given the central role of chromatin in mediating regulatory signals and controlling DNA access, and the paucity of recognizable sequence signals. Specific histone modifications correlate with regulator binding, transcriptional initiation and elongation, enhancer activity and repression^{1,3–6}. Combinations of modifications can provide even more precise insight into chromatin state^{7,8}.

Here we apply a high-throughput pipeline to map nine chromatin marks and input controls across nine cell types. We use recurrent combinations of marks to define 15 chromatin states corresponding to repressed, poised and active promoters, strong and weak enhancers, putative insulators, transcribed regions, and large-scale repressed and inactive domains. We use directed experiments to validate biochemical and functional distinctions between states.

The resulting chromatin state maps portray a highly dynamic landscape, with the specific patterns of change across cell types revealing strong correlations between interacting functional elements. We use correlated patterns of activity between chromatin state, gene expression and regulator activity to connect enhancers to likely target genes, to predict cell-type-specific activators and repressors, and to identify individual binding motifs responsible for these interactions.

Our results have implications for the interpretation of genome-wide association studies (GWASs). We find that disease variants frequently coincide with enhancer elements specific to a relevant cell

type. In several cases, we can predict upstream regulators whose regulatory motif instances are affected or target genes whose expression may be altered, thereby suggesting specific mechanistic hypotheses for how disease-associated genotypes lead to the observed disease phenotypes.

Results

Systematic mapping of chromatin marks in multiple cell types

To explore chromatin state in a uniform way across multiple cell types, we applied a production pipeline for chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) to generate genome-wide chromatin data sets (Methods and Fig. 1a). We profiled nine human cell types, including common lines designated by the ENCODE consortium¹ and primary cell types. These consist of embryonic stem cells (H1 ES), erythrocytic leukaemia cells (K562), B-lymphoblastoid cells (GM12878), hepatocellular carcinoma cells (HepG2), umbilical vein endothelial cells (HUVEC), skeletal muscle myoblasts (HSMC), normal lung fibroblasts (NHLF), normal epidermal keratinocytes (NHEK) and mammary epithelial cells (HMEC).

We used antibodies for histone H3 lysine 4 trimethylation (H3K4me3), a modification associated with promoters^{4,5,9}; H3K4me2 (dimethylation), associated with promoters and enhancers^{1,3,6,9}; H3K4me1 (methylation), preferentially associated with enhancers^{1,6}; lysine 9 acetylation (H3K9ac) and H3K27ac, associated with active regulatory regions^{9,10}; H3K36me3 and H4K20me1, associated with transcribed regions^{3–5}; H3K27me3, associated with Polycomb-repressed regions^{3,4}; and CTCF, a sequence-specific insulator protein with diverse functions¹¹. We validated each antibody by western blots and peptide competitions, and sequenced input controls for each cell type. We also collected data for H3K9me3, RNA polymerase II (RNAPII) and H2A.Z (also known as H2AFZ) in a subset of cells.

¹Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ²MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, Massachusetts 02139, USA. ³Howard Hughes Medical Institute, Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. ⁴Center for Systems Biology and Center for Cancer Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

*These authors contributed equally to this work.

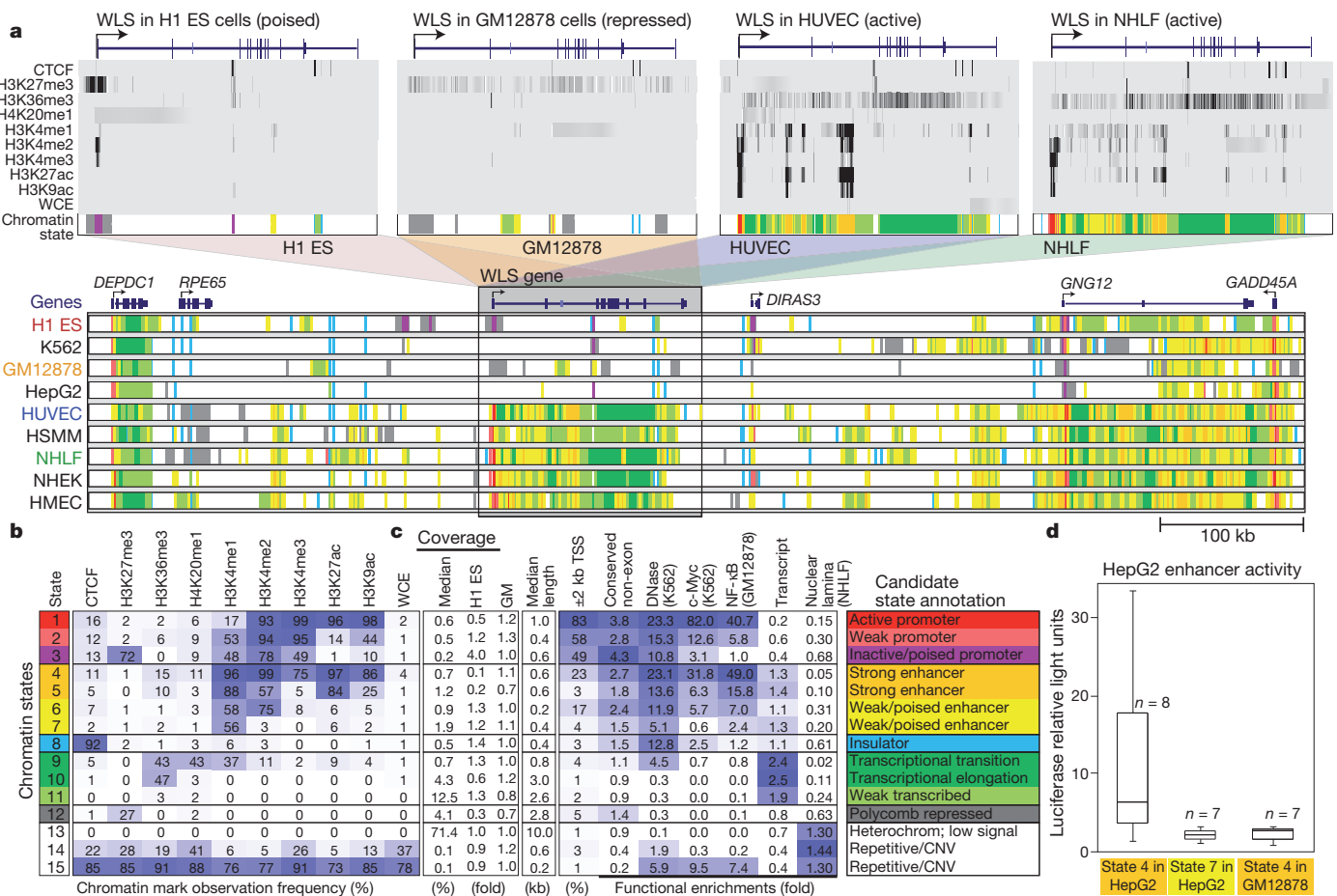


Figure 1 | Chromatin state discovery and characterization. **a**, Top: profiles for nine chromatin marks (greyscale) are shown across the WLS gene in four cell types, and summarized in a single chromatin state annotation track for each (coloured according to **b**). WLS is poised in ESCs, repressed in GM12878 and transcribed in HUVEC and NHLF. Its TSS switches accordingly between poised (purple), repressed (grey) and active (red) promoter states; enhancer regions within the gene body become activated (orange, yellow); and its gene body changes from low signal (white) to transcribed (green). These chromatin state changes summarize coordinated changes in many chromatin marks; for example, H3K27me3, H3K4me3 and H3K4me2 jointly mark a poised promoter, whereas loss of H3K27me3 and gain of H3K27ac and H3K9ac mark promoter activation. WCE, whole-cell extract. Bottom: nine chromatin state tracks, one per cell type, in a 900-kb region centred at WLS, summarizing 90 chromatin tracks in directly interpretable dynamic annotations and showing activation and repression patterns for six genes and hundreds of regulatory regions, including enhancer states. **b**, Chromatin states learned jointly across

This resulted in 90 chromatin maps corresponding to ~2,400,000,000 reads covering ~100,000,000,000 bases across nine cell types, which we set out to interpret computationally.

Learning a common set of chromatin states across cell types

To summarize these data sets into nine readily interpretable annotations, one per cell type, we applied a multivariate hidden Markov model that uses combinatorial patterns of chromatin marks to distinguish chromatin states⁸. The approach explicitly models mark combinations in a set of ‘emission’ parameters and spatial relationships between neighbouring genomic segments in a set of ‘transition’ parameters (Methods). It has the advantage of capturing regulatory elements with greater reliability, robustness and precision than is possible by studying individual marks⁸.

We learned chromatin states jointly by creating a virtual concatenation of all chromosomes from all cell types. We selected 15 states that showed distinct biological enrichments and were consistently recovered (Fig. 1a, b and Supplementary Fig. 1). Even though states

cell types by a multivariate hidden Markov model. The table shows emission parameters learned *de novo* on the basis of genome-wide recurrent combinations of chromatin marks. Each entry denotes the frequency with which a given mark is found at genomic positions corresponding to the chromatin state. **c**, Genome coverage, functional enrichments and candidate annotations for each chromatin state. Blue shading indicates intensity, scaled by column. CNV, copy number variation; GM, GM12878. **d**, Box plots depicting enhancer activity for predicted regulatory elements. Sequences 250 bp long corresponding either to strong or weak/poised HepG2 enhancer elements or to GM12878-specific strong enhancer elements were inserted upstream of a luciferase gene and transfected into HepG2. Reporter activity was measured in relative light units. Robust activity is seen for strong enhancers in the matched cell type, but not for weak/poised enhancers or for strong enhancers specific to a different cell type. Boxes indicate 25th, 50th and 75th percentiles, and whiskers indicate 5th and 95th percentiles.

were learned *de novo* solely on the basis of the patterns of chromatin marks and their spatial relationships, they showed distinct associations with transcriptional start sites (TSSs), transcripts, evolutionarily conserved non-coding regions, DNase hypersensitive sites¹², binding sites for the regulators c-Myc¹³ (MYC) and NF-κB¹⁴, and inactive genomic regions associated with the nuclear lamina¹⁵ (Fig. 1c).

We distinguished six broad classes of chromatin states, which we refer to as promoter, enhancer, insulator, transcribed, repressed and inactive states (Fig. 1c). Within them, active, weak and poised⁴ promoters (states 1–3) differ in expression level, strong and weak candidate enhancers (states 4–7) differ in expression of proximal genes, and strongly and weakly transcribed regions (states 9–11) also differ in their positional enrichments along transcripts. Similarly, Polycomb-repressed regions (state 12) differ from heterochromatic and repetitive states (states 13–15), which are also enriched for H3K9me3 (Supplementary Figs 2–4).

The states vary widely in their average segment length (~500 base pairs (bp) for promoter and enhancer states versus 10 kb for inactive

regions) and in the portion of the genome covered ($<1\%$ for promoter and enhancer states versus $>70\%$ for inactive state 13). For each state, coverage was relatively stable across cell types (Supplementary Fig. 5), with the exception of embryonic stem cells (ESCs) in which the poised promoter state is more abundant but strong enhancer and Polycomb-repressed states are depleted, consistent with the unique biology of pluripotent cells^{4,16}.

We confirmed that promoter and enhancer states showed distinct biochemical properties (Supplementary Fig. 6). RNAPII was highly enriched at strong promoters, weakly enriched at strong enhancers and nearly undetectable at weak or poised enhancers, consistent with strong transcription at promoters and reports of weak transcription at active enhancers^{17,18}. H2A.Z, a histone variant associated with nucleosome-free regions¹⁹, was enriched in active promoters and strong enhancers, consistent with nucleosome displacement at TSSs and sites of abundant transcription factor binding in active enhancers.

We also used luciferase reporter assays to validate the functionality of predicted enhancers, the distinction between strong and weak enhancer states, and their predicted cell type specificity. We tested strong enhancers, weak enhancers and strong enhancers specific to an unmatched cell type by transfection in HepG2. We observed strong luciferase activity only for strong enhancer elements from the matched cell type (Fig. 1d).

These results and additional properties of the model (Supplementary Figs 7–10) suggest that chromatin states are an inherent, biologically informative feature of the genome. The framework enables us to reason about coordinated differences in marks by directly studying chromatin state changes between cell types (which we refer to as ‘changes’ or ‘dynamics’ without implying any temporal relationship).

Extent and significance of chromatin state changes across cell types

We next explored the extent to which chromatin states vary between pairs of cell types. The overall patterns of variability (Supplementary Figs 11 and 12) suggest that regulatory regions vary drastically in activity level across cell types. Enhancer states show frequent interchange between strong and weak, and promoter states vary between active, weak and poised. Promoter states seem more stable than enhancers; they are eight times more likely to remain promoter states, controlling for coverage. Switching was also observed among promoter, enhancer and transcriptional transition states, but no preferential changes to other groups were found. These general patterns suggest that despite varying activity levels, enhancer and promoter regions tend to preserve their chromatin identity as regions of regulatory potential.

Chromatin state differences between cell types relate to cell-type-specific gene functions. An unbiased clustering of chromatin state profiles across annotated TSSs in lymphoblastoid and skeletal muscle cells distinguished informative patterns predictive of downstream gene expression and functional gene classes (Supplementary Figs 13 and 14). Cell-type-specific patterns were also evident when TSSs were simply assigned to the most prevalent chromatin state. Promoters active in skeletal muscle were associated with extracellular structure genes (8.5-fold enrichment), those active in lymphoblastoid cells were associated with immune response genes (7.2-fold enrichment) and those active in both were associated with metabolic housekeeping genes.

Clustering of promoter and enhancer states on the basis of their activity patterns

Extending our pairwise promoter analysis, we clustered active promoter and strong enhancer regions across all cell types (Methods). This revealed clusters showing common activity and associated with highly coherent functions (Fig. 2). For promoter clusters, these include immune response (GM12878-specific clusters, $P < 10^{-18}$), cholesterol transport (HepG2 specific, $P < 10^{-4}$) and metabolic processes (all cells, $P < 10^{-13}$). Remarkably, genes assigned to enhancer clusters by proximity also showed strong functional enrichments, including immune

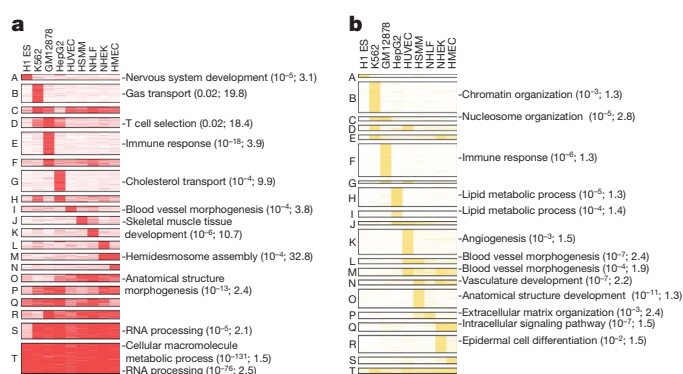


Figure 2 | Cell-type-specific promoter and enhancer states and associated functional enrichments. **a**, Clustering of genomic locations (rows) assigned to active promoter state 1 (red) across cell types (columns) reveals 20 common patterns of activity (A–T; Methods). For each cluster, enriched gene ontology terms are shown with hypergeometric *P* value and fold enrichment, based on the nearest TSS. For most clusters, several cell types show strong (dark red) or moderate (light red) activity. **b**, Analogous clustering and functional enrichments for strong enhancer state 4 (yellow). Enhancer states show greater cell type specificity, with most clusters active in only one cell type.

response (GM12878 specific, $P < 10^{-6}$), lipid metabolism (HepG2 specific, $P < 10^{-5}$) and angiogenesis (HUVEC specific, $P < 10^{-3}$).

Promoters and enhancers differed in their overall specificities. The majority of promoter clusters showed activity in multiple cell types, consistent with previous work^{5,10} (Fig. 2a). Enhancer clusters are significantly more cell type specific, with few regions showing activity in more than two cell types and a majority being specific to a single cell type (Fig. 2b).

We also found differences in the relative contributions of enhancer-based and promoter-based regulation among gene classes. Developmental genes seem to be strongly regulated by both, showing the highest number of proximal enhancers and diverse promoter states, including poised and Polycomb repressed (Supplementary Fig. 15). Tissue-specific genes (for example immune genes and steroid metabolism genes) seem to be more dependent on enhancer regulation, showing multiple tissue-specific enhancers but less diverse promoter states. Lastly, housekeeping genes are primarily promoter regulated, with few enhancers in their vicinities.

Overall, this dynamic view of the chromatin landscape suggests that multicell chromatin profiles can be as productive for systems biology as expression analysis has traditionally been, and may hold additional information on genome regulatory programs, which we explore next.

Correlations in activity profiles link enhancers to target genes

We next investigated functional interconnections among enhancers, the factors that activate or repress them, and the genes whose expression they regulate, by defining ‘activity profiles’ for each across the cell types (Fig. 3). We complemented these enhancer activity profiles (Fig. 3a) with profiles for gene expression (Fig. 3b), sequence motif enrichment (Fig. 3d) and the expression of transcription factors recognizing each motif (Fig. 3e). We used correlations between these profiles to probabilistically link enhancers to their downstream targets and upstream regulators (Methods).

We found that patterns of enhancer activity (Figs 2b and 3a) correlated strongly with patterns of nearest-gene expression (Fig. 3b; correlation, >0.9 in 16 of 20 clusters). Because this correlation remained high even for large distances (>50 kb), we used activity correlation as a complement to genomic distance for linking enhancers to target genes (Methods). Activity-based linking yielded an increase in functional gene class enrichment for several clusters (Supplementary Fig. 16).

We validated our approach using quantitative trait locus mapping studies that use covariation between single nucleotide polymorphism (SNP) alleles and gene expression levels to link *cis*-regulatory regions

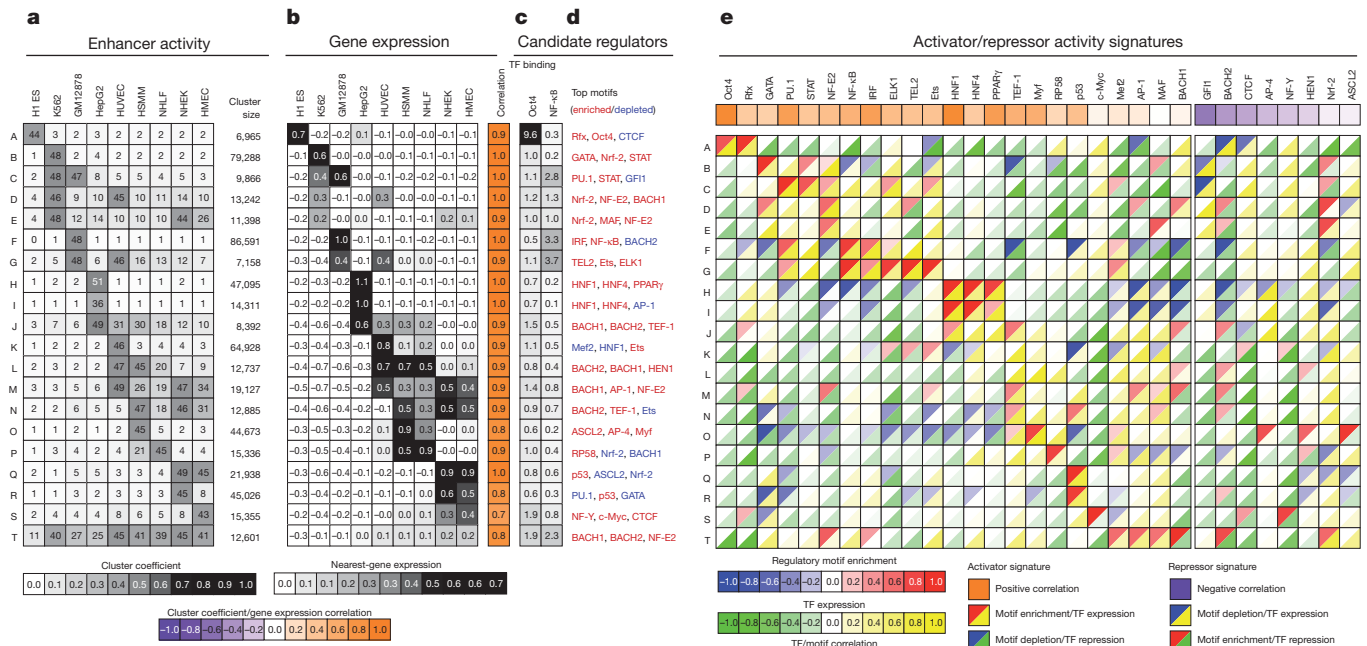


Figure 3 | Correlations in activity patterns link enhancers to gene targets and upstream regulators. **a**, Average enhancer activity across the cell types (columns) for each enhancer cluster (rows) defined in Fig. 2b (labelled A–T) and number of 200-bp windows in each cluster. **b**, Average messenger RNA expression of nearest gene across the cell types and correlation with enhancer activity profile from **a**. High correlations between enhancer activity and gene expression provide a means of linking enhancers to target genes. **c**, Enrichment for Oct4 binding in ESCs²⁴ and NF-κB binding in lymphoblastoid cells¹⁴ for each cluster. TF, transcription factor. **d**, Strongly enriched (red) or depleted (blue) motifs for each cluster, from a catalogue of 323 consensus motifs. Rfx: Rfx family; Nrf-2: NFE2L2; STAT: STAT family; Ets: Ets family; Mef2: MEF2A and MYEF2;

to target genes. Investigation of four recent quantitative trait locus studies in liver²⁰ and lymphoblastoid cells^{21–23} revealed remarkable agreement with our enhancer predictions. Enhancers linked to a given target gene by our method were significantly enriched for SNPs correlated with the gene's expression level (Supplementary Fig. 17), thus confirming our enhancer–gene linkages with orthogonal data.

Correlations with transcription factor expression and motif enrichment predict upstream regulators

We next predicted, on the basis of regulatory motif enrichments, sequence-specific transcription factors likely to target enhancers in a given cluster. This implicated a number of transcription factors whose known biological roles matched the respective cell types (Fig. 3d and Supplementary Fig. 18). When ChIP-seq data on the relevant cell type was available, we confirmed that enriched motifs were preferentially bound by the cognate factor (Fig. 3c). Oct4 (POU5F1) motif instances in cluster A (ESC-specific enhancers) were preferentially bound by Oct4 in ESCs²⁴, and NF-κB motif instances in cluster F (lymphoblastoid-specific enhancers) were preferentially bound by NF-κB in lymphoblastoid cells¹⁴. In both cases, motif instances in cell-type-specific enhancers showed a ~5-fold increase in binding in comparison with other enhancers.

However, sequence-based motif enrichments do not distinguish causality. Enrichment could reflect a parallel binding event that does not affect the chromatin state, or the motif could actually be antagonistic to the enhancer state through specific repression in orthogonal cell types. To distinguish between these possibilities, we complemented the motif enrichments with cell-type-specific expression for the corresponding transcription factors (Fig. 3e). We then correlated a 'motif score' based on motif enrichment in a given cluster, and a 'transcription factor expression score' based on the agreement between

Myf: Myf family; NF-Y: NFYA, NFYB and NFYC. **e**, Predicted causal regulators for each cluster based on positive (activators) or negative (repressors) correlations between motif enrichment (top left triangles) and transcription factor expression (bottom right triangles). For example, the red–yellow combination indicates that Oct4 is a positive regulator of ESC-specific enhancers, as its motif-based predicted targets are enriched (red upper triangle) for enhancers active in ESCs (cluster A), and the Oct4 gene is expressed specifically in ESCs, resulting in a positive transcription factor expression correlation (yellow triangle). Overall correlations between motif enrichment and transcription factor expression across all clusters denote predicted activators (positive correlation, orange) and repressors (negative correlation, purple).

the transcription factor expression pattern and the cluster activity profile (Methods). A positive correlation between the two scores implies that the transcription factor may be establishing or reinforcing the chromatin state. A negative correlation would instead imply that the transcription factor may act as a repressor. For example, in addition to the enrichment of the Oct4 motif in the ESC-specific cluster A, Oct4 is specifically expressed in ESCs, leading to the prediction that it is a causal regulator of ESCs (Fig. 3e), consistent with known biology¹⁶.

For 18 of the 20 clusters, this analysis revealed one or more candidate regulators. Recovery of known roles for well-studied regulators validated our approach. For example, HNF1 (HNF1A), HNF4 (HNF4A) and PPARγ (PPARG) are predicted as activators of HepG2-specific enhancers (clusters H and I), PU.1 (SPI1) and NF-κB as activators of lymphoblastoid (GM12878) enhancers (clusters C, F and G), GATA1 as an activator of K562-specific enhancers (cluster B) and Myf family members as HSMM enhancers^{14,25–27} (cluster O).

The analysis also revealed potentially novel regulatory interactions. ETS-related factors (ELK1, TEL2 (ETV7) and Ets family members) are predicted activators of enhancers active in both GM12878 and HUVEC (cluster G) but not of GM12878-specific or HUVEC-specific clusters, emphasizing the value of unbiased clustering. These connections are consistent with reported roles for ETS factors in lymphopoiesis and endothelium²⁸. The prediction of p53 (TP53) as an activator in HSMM, NHLF, NHEK and HMEC (clusters N, Q and R) probably reflects its maintained activity in these primary cells, as opposed to cell models in which it may be suppressed by mutation (K562)²⁹, viral inactivation (GM12878)³⁰ or cytoplasmic localization (ESCs)³¹. A widespread role for p53 in regulating distal elements is consistent with its known binding to distal regions^{32,33}.

Our analysis also revealed several repressor signatures, including GF11 in K562 and GM12878 (clusters B and C) and BACH2 in ESCs

(cluster A). Both regulators are known to repress transcription by recruiting histone deacetylases and methyltransferases to proximal promoters^{34,35}, and GFI1 has also been implicated in the silencing of satellite repeats³⁵. Our regulatory inferences suggest that these regulators also modulate chromatin to inhibit enhancer activity, thus suggesting a new mechanism for distal gene regulation.

Validation of predicted binding events and regulatory outcomes

The regulatory inferences above imply transcription-factor-binding events at motif instances within enhancer regions in specific cellular contexts, and we sought to validate these inferences using a general molecular signature. Binding events are associated with nucleosome displacement, a structural change evident in ChIP-seq data for histones³⁶. We thus studied local depletions in the chromatin intensity profiles ('dips') as these are indicative of transcription factor binding. We confirmed that dips were present in individual signal tracks for active enhancers and were associated with preferential sequence conservation and regulatory motif instances (Fig. 4a).

To test our specific predictions, we superimposed chromatin profiles of coordinately regulated enhancer regions, anchoring them on the implied motif instances. Striking dips precisely coincide with regulatory motifs, and are both cell type specific and region specific, exactly as predicted (Fig. 4b, c). Because dips only appear when the factor is expressed, they also support the identity of the *trans*-acting transcription factor.

To confirm that predicted causal motifs contribute to enhancer activity, we used luciferase reporters. Our model implicated HNF regulators as activators of HepG2-specific enhancers (Fig. 3), and context-specific dips supported binding interactions (Fig. 4c). We thus selected for functional analysis ten sites with HNF motifs showing dips in strong HepG2-specific enhancers, and evaluated them with and without the HNF motif. We found that permutation of

the motif consistently led to a reduction in enhancer activity (Fig. 4d), supporting its predicted causal role.

Assigning candidate regulatory functions to disease-associated variants

Finally, we explored whether our chromatin annotations and regulatory predictions can provide insight into sequence variants associated with disease phenotypes. To that effect, we gathered a large set of non-coding SNPs from GWAS catalogues, an exceedingly small proportion of which are understood at present³⁷.

We found that disease-associated SNPs are significantly more likely to coincide with strong enhancers (states 4 and 5; twofold enrichment, $P < 10^{-10}$), despite the fact that no notable association with these states are seen for SNPs in general or for those SNPs tested in the studies. To test whether SNPs associated with a particular disease might have even more specific correspondences, we examined 426 GWAS data sets. We identified ten studies^{38–47} whose variants showed significant correspondences to cell-type-specific strong enhancer states (Methods and Fig. 5a).

Individual variants from these studies were strongly enriched in enhancer states specifically active in relevant cell types (Fig. 5a, b). For example, SNPs associated with erythrocyte phenotypes³⁸ were found in erythrocytic leukaemia cell (K562) enhancers, SNPs associated with systemic lupus erythematosus³⁹ were found in lymphoblastoid cell (GM12878) enhancers and SNPs associated with triglyceride⁴⁰ phenotypes or blood lipid phenotypes⁴¹ were found in hepatocellular carcinoma cell (HepG2) enhancers. We also applied our model to chromatin data for T cells³ (Supplementary Fig. 19), for which strong enhancer states correlated to variants associated with risk of childhood acute lymphoblastic leukaemia⁴⁸, further validating our approach.

We also used our predicted enhancer/target gene associations to find candidate downstream genes whose expression might be affected by *cis* changes occurring in the enhancer region. Although most of the predicted target genes are proximal to the enhancer, a subset of more distal predicted targets could reflect novel candidates for the disease phenotypes (Fig. 5b).

In addition, we identified several instances in which a lead GWAS variant does not correspond to a particular chromatin element but a linked variant coincides with an enhancer with the predicted cell type specificity (Fig. 5c). Thus, chromatin profiles may provide a general means of triaging variants within a haplotype block, a common problem faced in GWASs.

Lastly, we identified several cases in which a disease-associated SNP created or disrupted a regulatory motif instance for a predicted causal transcription factor in the relevant cell type (Fig. 5d), suggesting a specific molecular mechanism by which the disease-associated genotype could lead to the observed disease phenotype consistent with our regulatory predictions.

Discussion

Our work demonstrates the power of multicell chromatin profiles as an additional and dynamic layer of genome annotation. We presented methods to distinguish different classes of functional elements, elucidate their cell type specificities and predict *cis*-regulatory interactions that drive gene expression programs. By intersecting our predictions with non-coding SNPs from GWAS data sets, we proposed potential mechanistic explanations for disease variants, either through their presence within cell-type-specific enhancer states or by their effect on binding motifs for predicted regulators.

Chromatin states drastically reduced the large combinatorial space of 90 chromatin data sets (2^{90} combinations) to a manageable set of biologically interpretable annotations, thus providing an efficient and robust way to track coordinated changes across cell types. This allowed the systematic identification and comparison of more than 100,000 promoter and enhancer elements. Both types of element are cell type specific, are associated with motif enrichments and assume

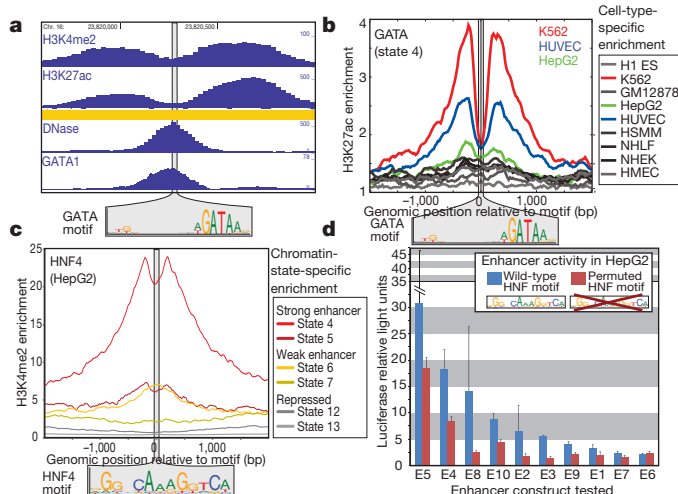


Figure 4 | Validation of regulatory predictions by nucleosome depletions and enhancer activity. **a**, Dips in chromatin intensity profiles in a K562-specific strong enhancer (orange) coincide with a predicted causal GATA motif instance (logo). The dips probably reflect nucleosome displacement associated with transcription factor binding, supported by DNase hypersensitivity¹² and GATA1 binding²⁵. **b**, Superposition of H3K27ac signal across loci containing GATA motifs, centred on motif instances, shows dips in K562, as predicted. **c**, Superposition of H3K4me2 signal for HepG2 shows dips over HNF4 motifs in strong enhancer states, as predicted. **d**, HepG2-specific strong enhancers with predicted causal HNF motifs were tested in reporter assays. Constructs with permuted HNF motifs (red) led to significantly reduced luciferase activity in comparison with wild type (blue), with an average twofold reduction. Data shown are mean luciferase relative light units over three replicates and 95% confidence intervals.

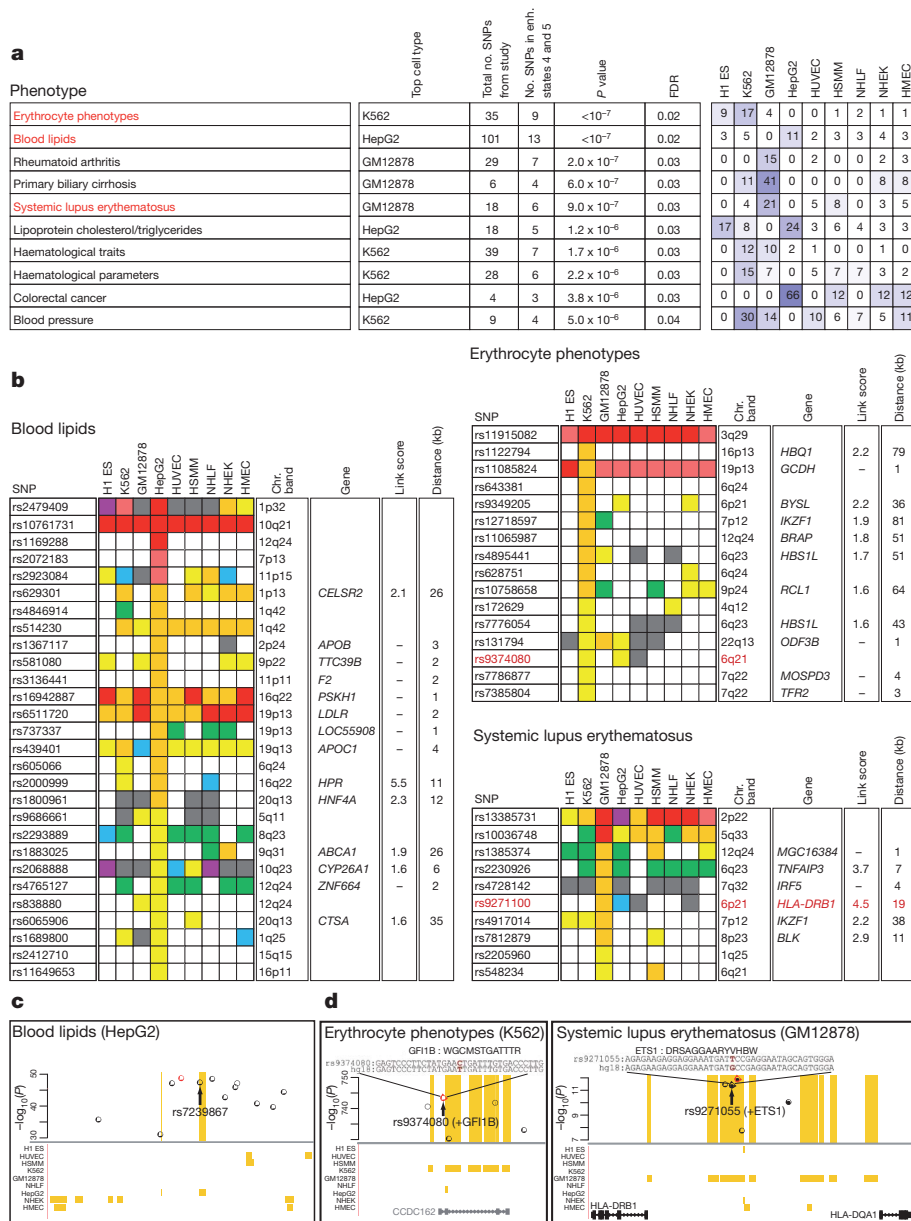


Figure 5 | Disease variants annotated by chromatin dynamics and regulatory predictions.
a, Intersection of strong enhancer states (4 and 5) with disease-associated SNPs from GWASs shows significant enrichment (blue shading) in relevant cell types (Methods). Fold enrichments of the SNPs in strong enhancer states for each cell type are indicated. FDR, false-discovery rate. **b**, For three GWAS data sets^{38–40}, state annotations are shown for a subset of lead SNPs in the nine cell types (colours as in Fig. 1b, except state 11 is white). The strong enhancer state (orange) is most prevalent in cell types related to the phenotype. For SNPs overlapping strong enhancers, proximal genes with correlated expression are indicated, with linking score and distance. **c**, Example GWAS locus with blood lipid trait⁴¹ association, where the lead variant (red circle) has no functional annotation but a linked SNP (arrow) coincides with a HepG2-specific strong enhancer (orange) and may represent a causal variant. Strong enhancer annotations are shown for all cell types. **d**, Example GWAS loci where a disease SNP affects a conserved instance of a predicted causal motif. Left: lead SNP rs9374080 in the erythrocyte phenotype GWAS³⁸ is <100 bp from a strong enhancer in K562 and strengthens a motif for GFI1B, a predicted repressor in K562 (Fig. 3e). Right: SNP rs9271055 associated with lupus³⁹ coincides with a lymphoblastoid (GM12878) strong enhancer and strengthens a motif for ETS1, a predicted activator of lymphoblastoid enhancers (Fig. 3e). This factor is further implicated by lupus-associated variants that directly affect the *ETS1* locus³⁹.

strong, weak and poised states that correlate with neighbouring gene expression and function. Enhancers showed very high tissue specificity, enrichment in the vicinity of developmental and cell-type-specific genes, and predictive power for proximal gene expression, reinforcing their roles as sentinels of tissue-specific gene expression⁴⁹. By elucidating enhancers systematically, and linking them to upstream regulators and downstream genes, our analysis can help provide a missing link between regulators and target genes. The power of the approach should increase considerably as additional phenotypically distinct cell types are surveyed, and should enable a greater proportion of enhancer elements to be incorporated into the connectivity network.

The inferred *cis*-regulatory interactions make specific testable predictions, many of which were confirmed through additional experiments and analyses. Our enhancer/target gene linkages are supported by *cis*-regulatory inferences from quantitative trait locus mapping studies. Predicted transcription factor/motif interactions within cell-type-specific enhancers were confirmed in specific cases by transcription factor binding and more generally by depletions in the chromatin profiles at causal motifs in appropriate cellular contexts. Motifs predicted as causal regulators of cell-type-specific enhancers were also confirmed in enhancer assays.

The regulatory inferences afforded by multicell chromatin profiles are unique and highly complementary to data sets for transcription factor binding, expression, chromatin accessibility, nucleosome positioning and chromosome conformation⁵⁰. For example, our regulatory predictions can help focus the spectrum of transcription-factor-binding events to a smaller number of functional interactions. The 'chromatin-centric' approach also complements the extensive body of work on biological network inference from expression data with the potential to introduce enhancers and other genomic elements into connectivity networks.

Our study has important implications for the understanding of disease. Our detailed and dynamic functional annotations of the relatively uncharted non-coding genome can facilitate the interpretation of GWAS data sets by predicting specific cell types and regulators related to specific diseases and phenotypes. Furthermore, the connections derived for enhancer regions, to upstream regulators and downstream genes, suggest *cis*- and *trans*-acting interactions that may be modulated by the sequence variants. Although the present study represents only a first, small step in this direction, we expect that future iterations with a greater diversity of cell types and improved methodologies will help define the molecular underpinnings of human disease.

METHODS SUMMARY

We performed ChIP-seq analysis in biological replicate as previously described⁴, using antibodies validated by western blots and peptide competitions. ChIP DNA and input controls were sequenced using the Illumina Genome Analyser. Expression profiles were acquired using Affymetrix GeneChip arrays. Chromatin states were learned jointly by applying a hidden Markov model⁸ to ten data tracks for each of the nine cell types. We focused on a 15-state model that provides sufficient resolution to resolve biologically meaningful patterns yet is reproducible across cell types when independently processed. We used this model to produce nine genome-wide chromatin state annotations, which were validated by additional ChIP experiments and reporter assays. Multicell type clustering was conducted on locations assigned to strong promoter state 1 (or strong enhancer state 4) in at least one cell type using the *k*-means algorithm. We predicted enhancer/target gene linkages by correlating normalized signal intensities of H3K27ac, H3K4me1 and H3K4me2 with gene expression across cell types as a function of distance to the TSS. Upstream regulators were predicted using a set of known transcription factor motifs assembled from multiple sources. Motif instances were identified by sequence match and evolutionary conservation. We based *P* values for GWAS studies on randomizing the location of SNPs, and based the false-discovery rate on randomizing the assignment of SNPs across studies. Data sets are available from the ENCODE website (<http://genome.ucsc.edu/ENCODE>), the supporting website for this paper (http://compbio.mit.edu/ENCODE_chromatin_states) and the Gene Expression Omnibus (GSE26386).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 August 2010; accepted 4 February 2011.

Published online 23 March 2011.

- Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
- Kim, H. D., Shay, T., O'Shea, E. K. & Regev, A. Transcriptional regulatory circuits: predicting numbers from alphabets. *Science* **325**, 429–432 (2009).
- Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
- Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
- Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
- Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39**, 311–318 (2007).
- Hon, G., Wang, W. & Ren, B. Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput. Biol.* **5**, e1000566 (2009).
- Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature Biotechnol.* **28**, 817–825 (2010).
- Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181 (2005).
- Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
- Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211 (2009).
- Hansen, R. S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl Acad. Sci. USA* **107**, 139–144 (2010).
- Raha, D. *et al.* Close association of RNA polymerase II and many transcription factors with Pol III genes. *Proc. Natl Acad. Sci. USA* **107**, 3639–3644 (2010).
- Kasowski, M. *et al.* Variation in transcription factor binding among humans. *Science* **328**, 232–235 (2010).
- Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
- Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567–582 (2008).
- De Santa, F. *et al.* A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, e1000384 (2010).
- Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
- Talbert, P. B. & Henikoff, S. Histone variants — ancient wrap artists of the epigenome. *Nature Rev. Mol. Cell Biol.* **11**, 264–275 (2010).
- Schadt, E. E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).
- Pickrell, J. K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768–772 (2010).
- Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773–777 (2010).
- Veyrieras, J. B. *et al.* High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* **4**, e1000214 (2008).
- Kunaro, G. *et al.* Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genet.* **42**, 631–634 (2010).
- Fujiwara, T. *et al.* Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. *Mol. Cell* **36**, 667–681 (2009).
- Lemaigre, F. & Zaret, K. S. Liver development update: new embryo models, cell lineage control, and morphogenesis. *Curr. Opin. Genet. Dev.* **14**, 582–590 (2004).
- Sabourin, L. A. & Rudnicki, M. A. The molecular regulation of myogenesis. *Clin. Genet.* **57**, 16–25 (2000).
- Bartel, F. O., Higuchi, T. & Spyropoulos, D. D. Mouse models in the study of the Ets family of transcription factors. *Oncogene* **19**, 6443–6454 (2000).
- Law, J. C., Ritke, M. K., Yalowich, J. C., Leder, G. H. & Ferrell, R. E. Mutational inactivation of the p53 gene in the human erythroid leukemic K562 cell line. *Leuk. Res.* **17**, 1045–1050 (1993).
- Forte, E. & Luftig, M. A. MDM2-dependent inhibition of p53 is required for Epstein-Barr virus B-cell growth transformation and infected-cell survival. *J. Virol.* **83**, 2491–2499 (2009).
- Solozobova, V., Rolletschek, A. & Blattner, C. Nuclear accumulation and activation of p53 in embryonic stem cells after DNA damage. *BMC Cell Biol.* **10**, 46 (2009).
- Cawley, S. *et al.* Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**, 499–509 (2004).
- Wei, C. L. *et al.* A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**, 207–219 (2006).
- Hoshino, H. *et al.* Co-repressor SMRT and class II histone deacetylases promote Bach2 nuclear retention and formation of nuclear foci that are responsible for local transcriptional repression. *J. Biochem.* **141**, 719–727 (2007).
- Vassen, L., Fiolka, K. & Moroy, T. Gfi1b alters histone methylation at target gene promoters and sites of gamma-satellite containing heterochromatin. *EMBO J.* **25**, 2409–2419 (2006).
- He, H. H. *et al.* Nucleosome dynamics define transcriptional enhancers. *Nature Genet.* **42**, 343–347 (2010).
- Hindorf, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106**, 9362–9367 (2009).
- Ganesh, S. K. *et al.* Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nature Genet.* **41**, 1191–1198 (2009).
- Han, J. W. *et al.* Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nature Genet.* **41**, 1234–1237 (2009).
- Kathiresan, S. *et al.* Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nature Genet.* **40**, 189–197 (2008).
- Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
- Houlston, R. S. *et al.* Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nature Genet.* **40**, 1426–1435 (2008).
- Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nature Genet.* **41**, 666–676 (2009).
- Stahl, E. A. *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nature Genet.* **42**, 508–514 (2010).
- Liu, X. *et al.* Genome-wide meta-analyses identify three loci associated with primary biliary cirrhosis. *Nature Genet.* **42**, 658–660 (2010).
- Kamatani, Y. *et al.* Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nature Genet.* **42**, 210–215 (2010).
- Soranzo, N. *et al.* A genome-wide meta-analysis identifies 22 loci associated with eight hematology parameters in the HaemGen consortium. *Nature Genet.* **41**, 1182–1190 (2009).
- Papaemmanuil, E. *et al.* Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nature Genet.* **41**, 1006–1010 (2009).
- Visel, A., Rubin, E. M. & Pennacchio, L. A. Genomic views of distant-acting enhancers. *Nature* **461**, 199–205 (2009).
- Naumova, N. & Dekker, J. Integrating one-dimensional and three-dimensional maps of genomes. *J. Cell Sci.* **123**, 1979–1988 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the epigenomics community at the Broad Institute and the Bernstein and Kellis laboratories, and M. Daly, D. Altschuler and E. Lander for discussions and criticisms. We also thank M. Suva, E. Mendenhall and S. Gillespie for assistance with experiments, and L. Goff and A. Chess for critical reading of the manuscript. We acknowledge the Broad Institute Genome Sequencing Platform for their expertise and assistance with data production. This research was supported by the National Human Genome Research Institute under an ENCODE grant (U54 HG004570; B.E.B.), R01 HG004037 (M. Kellis), RC1 HG005334 (M. Kellis), the Howard Hughes Medical Institute (B.E.B.), the National Science Foundation (awards 0644282 (M. Kellis) and 0905968 (J.E.)) and the Sloan Foundation (M. Kellis).

Author Contributions J.E. conducted chromatin state analysis. J.E. and P.K. conducted regulatory motif analysis. J.E. and L.W. conducted GWAS SNP analysis. T.S.M., N.S. and T.D. implemented the ChIP-seq data processing pipeline. C.B.E., X.Z., L.W., R.I., M.C. and M. Ku developed the experimental pipeline and conducted experiments. M. Kellis designed and directed the computational analysis. B.E.B. designed the experimental approach and oversaw the work. J.E., M. Kellis and B.E.B. wrote the paper.

Author Information Sequencing and expression data has been deposited into the Gene Expression Omnibus under accession number GSE26386. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M. Kellis (manoli@mit.edu).

METHODS

Cell culture. Human H1 ES cells were cultured in TeSR media⁵¹ on Matrigel by Cellular Dynamics International. Cells were split with dispase and collected at a passage number between 30 and 40. Before collection, cells were karyotyped and stained for Oct4 to confirm pluripotency. K562 erythrocytic leukaemia cells (ATCC CCL-243, lot no. 4607240) were grown in suspension in RPMI medium (HyClone SH30022.02) with 10% fetal bovine serum (FBS) and 1% Antibiotic-Antimycotic (GIBCO 15240-062). Cell density was maintained at between 3×10^5 and 7×10^5 cells ml⁻¹. GM12878 B-lymphoblastoid cells (Coriell Cell Repositories, 'expansion A') were grown in suspension in RPMI 1640 medium with 15% FBS (not heat inactivated), 2 mM L-glutamine and 1% penicillin/streptomycin. Cells were seeded at a concentration of $\sim 2 \times 10^5$ viable cells ml⁻¹ with minimal disruption, and maintained at between 3×10^5 and 7×10^5 cells ml⁻¹. HepG2 hepatocellular carcinoma cells (ATCC HB-8065, lot no. 4968519) were cultured in DMEM (HyClone SH30022.02) with 10% FBS and 1% penicillin/streptomycin. Cells were trypsinized, resuspended to single-cell suspension, split to a confluence of between 15 and 20% and then collected at $\sim 75\%$ confluence. NHEK normal human epidermal keratinocytes isolated from skin (Lonza CC-2501, lot no. 4F1155J, passage 1) were grown in keratinocyte basal medium 2 (KGM-2 BulletKit, Lonza) supplemented with BPE, hEGF, hydrocortisone, GA-1000, transferrin, epinephrine and insulin. Cells were seeded at the recommended density (3,500 cells cm⁻²), subjected to two or three passages on polystyrene tissue culture plates and collected at a confluence of 70 to 80%. HSMM primary human skeletal muscle myoblasts (Lonza CC-2580, lot no. 6F4444, passage 2) were cultured in Smooth Muscle Growth Medium 2 (SkGM-2 BulletKit, Lonza) supplemented with rhEGF, dexamethasone, L-glutamine, FBS and GA-1000. Cells were seeded at the recommended density (3,500 cells cm⁻²), subjected to two or three passages and collected at a confluence of 50 to 70%. NHLF primary normal human lung fibroblasts (Lonza CC-2512, lot no. 4F0758, passage 2) were grown in Fibroblast Cell Basal Medium 2 (FGM-2 BulletKit, Lonza) supplemented with hFGF- β , insulin, FBS and GA-100. Cells were seeded at the recommended density (2,500 cells cm⁻²), subjected to two or three passages and collected at an approximate confluence of 80%. HUVEC primary human umbilical vein endothelial cells (Lonza CC-2517, lot no. 7F3239, passage 1) were grown in endothelial basal medium 2 (EBM-2 BulletKit, Lonza) supplemented with hFGF- β , hydrocortisone, VEGF, R3-IGF-1, ascorbic acid, heparin, FBS, hEGF and GA-1000. Cells were seeded at the recommended density (2,500–5,000 cells cm⁻²), subjected to two or three passages and collected at a confluence of 70 to 80%. HMEC primary human mammary epithelial cells from mammary reduction tissue (Lonza CC-2551, passage 7) were grown in mammary epithelia basal medium (MEGM BulletKit, Lonza) supplemented with hEGF- β , hydrocortisone, BPE, GA-1000 and insulin. Cells were seeded at the recommended density (2,500 cells cm⁻²), subjected to two or three passages and collected at 60 to 80% confluence.

Antibodies. ChIP assays were performed using the following antibody reagents: H3K4me1 (Abcam ab8895, lot 38311/659352), H3K4me2 (Abcam ab7766, lot 56293), H3K4me3 (Abcam ab8580, lot 331024; Millipore 04-473, lot DAM1623866), H3K9ac (Abcam ab44441, lot 455103/550799), H3K27ac (Abcam ab4729, lot 31456), H3K36me3 (Abcam ab9050, lot 136353), H4K20me1 (Abcam ab9051, lot 104513/519198), H3K27me3 (Millipore 07-449, lot DAM1387952/DAM1514011), CTCF (Millipore 07-729, lot 1350637), H3K9me3 (Abcam ab8898, lot 484088), H2A.Z (Millipore 07-594, lot DAM1504736) and RNAPII N terminus (Santa Cruz sc-899X, lot H0510). All antibody lots were extensively validated for specificity and efficacy in ChIP-seq. Western blots were used to confirm specific recognition of histone protein (or CTCF). Dot plots performed using arrayed histone tail peptides representing various modification states were used to confirm specificity for the appropriate modification. ChIP-seq assays performed on a common cell reagent were used to confirm consistency between different lots of the same antibody.

Chromatin immunoprecipitation. Cells were harvested by crosslinking with 1% formaldehyde in cell culture medium for 10 min at 37 °C. After quenching with the addition of 125 mM glycine for 5 min at 37 °C, the cells were washed twice with cold PBS containing protease inhibitor (Roche). After aspiration of all liquid, pellets consisting of $\sim 10^7$ cells were flash frozen and stored at -80 °C. Fixed cells were thawed and sonicated to obtain chromatin fragments of ~ 200 to 700 bp using a Bioruptor (Diagenode). Immunoprecipitation was performed as previously described, retaining a fraction of input 'whole-cell extract' as a control⁴. Briefly, sonicated chromatin was diluted tenfold and incubated with ~ 5 μ g antibody overnight. Antibody–chromatin complexes were pulled-down using protein A sepharose, washed and then eluted. After crosslink reversal and proteinase K treatment, immunoprecipitated DNA was extracted with phenol, precipitated in ethanol and treated with RNase. ChIP DNA was quantified by fluorometry using the Qubit assay (Invitrogen).

Next-generation sequencing. For each ChIP or control sample, ~ 5 ng of DNA was used to generate a standard Illumina sequencing library. Briefly, DNA fragments were end-repaired using the End-It DNA End-Repair Kit (Epicentre), extended with a 3' 'A' base using Klenow (3' \rightarrow 5' exo-, 0.3 U μ l⁻¹, NEB), ligated to standard Illumina adapters (75 bp with a 'T' overhang) using DNA ligase (0.05 U μ l⁻¹, NEB), gel-purified on 2% agarose, retaining products between 275 and 700 bp, and subjected to 18 PCR cycles. These libraries were quantified by fluorometry and evaluated by quantitative PCR or a multiplexed-digital-hybridization-based analysis⁵² (NanoString nCounter) to confirm representation and specific enrichment of DNA species. Libraries were sequenced in one or two lanes on the Illumina Genome Analyser using standard procedures for cluster amplification and sequencing by synthesis.

Expression profiling. Cytosolic RNA was isolated using RNeasy Columns (Qiagen) from the same cell lots as above. Gene expression profiles were acquired using Affymetrix GeneChip arrays. The data were normalized using the GenePattern expression data analysis package⁵³. CEL files were processed by RMA, quantile normalization and background correction. Two replicate expression data sets for each cell type were averaged and log₂-transformed. Gene-level normalization across cell types was computed by mean normalization.

Primary processing of sequencing reads. ChIP-seq reads were aligned to human genome build HG18 with MAQ (<http://maq.sourceforge.net/maq-man.shtml>) using default parameters. All reads were truncated to 36 bases before alignment. Signal density maps for visualization were derived by extending sequencing reads by 200 bp in the 3' direction (the estimated median size of ChIP fragments), and then counting the total number of overlapping reads at 25-bp intervals. Replicate ChIP-seq experiments were verified by comparing enriched intervals as previously described⁴, and were then combined into a single data set. For the hidden Markov model (HMM), density maps were derived by extending sequencing reads by 200 bp in the 3' direction and then assigning them to a single 200-bp window based on the midpoint of the extended read. These maps were then binarized at 200-bp resolution on the basis of a Poisson background model using a threshold of 10^{-4} .

Joint learning of HMM states across cell types. To handle data from the nine cell types, we concatenated their genomes to create an extended virtual genome that we used to train the HMM. We applied the model to ten tracks corresponding to the different chromatin marks and input using a multivariate HMM as previously described⁸. Here we used a Euclidean distance for determining initial parameters for the nested initialization step. After the HMM had learned and evaluated a set of roughly nested models, considering up to 25 states, we focused on a 15-state model that provides sufficient resolution to resolve biologically meaningful chromatin patterns and yet is highly reproducible across cell types when independently processed (Supplementary Fig. 7). We used this model to compute the probability that each location is in a given state, and then assigned each 200-bp interval to its most likely state for each cell type. Even though our model focuses on presence/absence frequencies of marks, we found that our states also capture signal intensity differences between high-frequency and low-frequency marks (Supplementary Fig. 9).

Enrichment analysis. For each state, enrichments for different annotations were computed at 200-bp resolution with the exception of conservation, which was computed at nucleotide resolution. We used annotations obtained through the UCSC Genome Browser⁵⁴ for RefSeq TSSs and transcribed regions⁵⁵, PhastCons⁵⁶, DNase-seq for K562 cells¹², c-Myc ChIP-seq for K562 cells¹³, NF- κ B ChIP-seq for GM12878¹⁴, Oct4 in ESCs²⁴ and nuclear lamina¹⁵. Gene functional group enrichments were determined using STEM⁵⁷ and biological process annotations in the Gene Ontology database⁵⁸. *P* values were calculated on the basis of the hypergeometric distribution and corrected for multiple testing using Bonferroni correction.

Comparing chromatin state assignments between cell types. For each pair of cell types, the chromatin state assignments at each genomic position were compared. We calculated the frequency with which each pair of states occurred, and normalized this against the expected frequency based on the amount of genome covered by each state. The fold enrichments in Fig. 2a reflect an aggregation across all 72 possible pairs of cell types.

Pairwise promoter clustering. Promoters for RefSeq genes were clustered on the basis of the most likely chromatin state assignment across a 2-kb region centred on the TSS. Clustering was performed jointly across GM12878 and HSMM, and was restricted to genes with corresponding Affymetrix expression. Briefly, each promoter was treated as a 330-element binary vector in which each component corresponded to a position along the promoter, cell type and state. Clustering was performed on these vectors using the *k*-means algorithm in MATLAB. Gene expression values were calculated on the basis of the corresponding Affymetrix probe set closest to the TSS.

Multicell type promoter and enhancer clustering. Promoter state clustering was performed for all 200-bp intervals assigned to the strong promoter state (state 1)

in at least one cell type. Each interval was represented by a single vector whose components are the estimated probabilities that it be in the strong promoter state for each of the nine cell types, accounting for model assignment uncertainty and biological noise. These were determined from the model posterior probabilities of state assignments and a comparison of state assignments in replicate experimental data. Clustering was performed using the *k*-means algorithm in MATLAB. We found that 20 clusters provided sufficient resolution to distinguish major cell-type-specific patterns. Enhancer state clustering was performed for all 200-bp intervals assigned to strong enhancer state 4 in at least one cell type using identical procedures. For the purposes of display in Fig. 2, the locations were randomly down-sampled. For the purpose of identifying enriched functional gene categories in Fig. 2b, enhancers were linked to the nearest TSS up to 50 kb distant excluding those within 5 kb. Enhancer–gene correspondences based on the nearest gene were used for the expression analysis of distance-based linked genes in Fig. 3b.

Linking enhancer locations to correlated genes. To predict linkages between enhancer states and target genes, we combined distance-based information with multicell type correlations between gene expression levels and normalized signal intensities for histone modifications associated with enhancer states (H3K4me1, H3K4me2 and H3K27ac). For each enhancer state (4–7), cell type, and 200-bp interval between 5 kb and 125 kb from the TSS, we trained logistic regression classifiers. The classifiers were trained to use mark intensity/expression correlation values to distinguish real instances of pairs of enhancer states and gene expression values from control pairs based on randomly re-assigning expression values to different genes. So that the classifiers learned a smooth and robust function at each position, we included as part of the training all enhancer state assignments within a 10-kb window centred at the position. The link score for a specific enhancer–gene linkage was defined as the ratio of the corresponding logistic regression classifier probability score to that for the randomized data.

For the evaluation of the expression quantitative trait loci (QTL) analysis, we used a link score threshold of 2.5. The expression QTL data was obtained from the University of Chicago QTL browser (<http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>). In the QTL evaluation, each SNP that overlapped a strong enhancer state (4 or 5), was within 125 kb of a TSS, excluding locations within 5 kb, and was associated with a gene for which we had expression data was considered eligible to be supported by our linked predictions. We computed the fraction we observed linked on the basis of our linked predictions relative to the fraction that would be expected to be linked conditioned on knowing the distance distributions of the SNPs relative to the gene TSS.

For the evaluation of linked predictions using the Gene Ontology database, we used the same link score threshold and compared gene assignments against the distance-based assignments defined above. The base set of genes in the enrichment analysis here were all genes that could be linked in at least one cluster.

Motif and transcription factor analysis. A database of known transcription factor motifs was collated by combining motifs from TRANSFAC⁵⁹ (version 11.3), JASPAR⁶⁰ (2010-05-07) and protein-binding microarray data sets^{61–63}. Motif instances in non-coding and non-repetitive regions of the genome were identified using these motifs and sequence conservation using a 29-way alignment of eutherian mammal genomes (K. Lindblad-Toh *et al.*, submitted). These were filtered using a significance threshold of $P < 4^{-8}$ for the motifs⁶⁴, and a confidence level based on conservation. Motifs were linked to corresponding transcription factors using metadata provided by the source. Motif enrichments for chromatin state clusters were computed as ratios to the instances of shuffled motifs, to correct for non-specific conservation and composition. A confidence interval was calculated for each ratio using Wilson score intervals ($z = 1.5$), selecting the most conservative value within the confidence interval. In cases where multiple motif variants were available for the same transcription factor, the one that showed the most variance in enrichment across clusters was selected.

For predicting causal activators and repressors, motif scores and transcription factor expression scores were correlated as follows. Motif scores were calculated as described above. Transcription factor expression scores were calculated for each cluster by correlating the expression of the transcription factor across the cell types with the activity profile of the enhancers in that cluster (defined by the cluster means from the *k*-means clustering). The motif scores and the transcription factor expression scores were then correlated against each other to identify positively and negatively correlated transcription factors.

Transcription factor/motif interactions predicted for strong enhancer states in specific cell types were validated by using the raw ChIP-seq tag enrichments as proxy for nucleosome positioning. For this purpose, sequencing reads were processed as above, except that the middle 75 bp of inferred ChIP fragments were used to derive signal density informative of nucleosome depletion (dips), as previously described³⁶. Superposition plots show tag enrichments relative to a uniform background computed on the basis of sequencing depth.

Quantitative real-time PCR. Enrichment ratios for RNAPII and H2A.Z ChIPs were determined relative to input chromatin by quantitative real-time PCR using an ABI 7900 detection system, in biological replicate as described previously⁶⁵. Regions used for validation correspond to three different chromatin states, including 13 for state 1 (arbitrarily selected), 11 for state 4 (arbitrarily selected but excluding regions within 2 kb of a state-1 annotation) and 11 for state 7 (arbitrarily selected but excluding regions within 2 kb of a state-1 or state-4 annotation). PCR primers are listed in Supplementary Data 1.

Functional enhancer assays. The SV40 promoter was first inserted between the HindIII and NcoI sites of pGL4.10 (Promega). Next, 250-bp sequences from the reference genome (hg18) corresponding to different chromatin states (eight from HepG2 state 4, seven from HepG2 state 7 and seven from GM12878 state 4) were synthesized (GenScript) and then inserted between the two SfiI sites upstream of the SV40 promoter. HepG2 cells were seeded into 96-well plates at a density of 5×10^4 cells per well and expanded overnight to ~50% confluency. The cells were then transfected with 400 ng of a pGL4.10-derived plasmid and 100 ng of pGL4.73 (Promega) using Lipofectamine LTX. Firefly and Renilla luciferase activities were measured 24 h post-transfection using Dual-Glow (Promega) and an EnVision 2103 multilabel reader (PerkinElmer), from triplicate experiments. Data are reported as light units relative to a control plasmid. For validation of causal transcription factor motifs, ten sequences of 250 bp corresponding to HepG2-specific strong enhancers (state 4) with dips and HNF motifs were tested as above, and compared with identical sequences except with the HNF motif permuted. Tested enhancer elements are listed in Supplementary Data 1.

GWAS SNP analysis. The GWAS variants and SNP coordinates were obtained from the NHGRI catalogue and the UCSC browser^{37,54} (October 30, 2010). This set was refined by extending the blood lipid GWAS⁴¹ set to contain all reported SNPs, and by bifurcating the haematological and biochemical traits study⁴⁶ into a haematological traits set and a biochemical traits set. We limited our analysis to studies reporting two or more associated SNPs. The variants from each study were intersected with chromatin states from each of the cell types. The reported *P* values were based on the overlap of associated SNPs with strong enhancer states 4 and 5. We controlled for non-independence between proximal SNPs by using a randomization test where SNPs were randomly shifted while preserving relative distance. We then defined an estimated false-discovery rate based on permutations in which SNPs were randomly re-assigned to different studies, and recomputed *P* values. Estimates of false-discovery rates based on these permutations control for multiple testing of studies and cell types and for general non-specific enrichments for states 4 and 5 with GWAS hits. Candidate gene targets were predicted for a subset of variants associated with enhancer states on the basis of the lead cell type using the linking method described above.

Data access. Data sets are available from the ENCODE website (<http://genome.ucsc.edu/ENCODE>), the supporting website for this paper (http://compbio.mit.edu/ENCODE_chromatin_states) and the Gene Expression Omnibus (GSE26386).

51. Ludwig, T. E. *et al.* Feeder-independent culture of human embryonic stem cells. *Nature Methods* **3**, 637–646 (2006).
52. Geiss, G. K. *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnol.* **26**, 317–325 (2008).
53. Reich, M. *et al.* GenePattern 2.0. *Nature Genet.* **38**, 500–501 (2006).
54. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
55. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61–D65 (2007).
56. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
57. Ernst, J. & Bar-Joseph, Z. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* **7**, 191 (2006).
58. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genet.* **25**, 25–29 (2000).
59. Matys, V. *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* **31**, 374–378 (2003).
60. Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **32**, D91–D94 (2004).
61. Berger, M. F. *et al.* Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**, 1266–1276 (2008).
62. Badis, G. *et al.* Diversity and complexity in DNA recognition by transcription factors. *Science* **324**, 1720–1723 (2009).
63. Berger, M. F. *et al.* Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nature Biotechnol.* **24**, 1429–1435 (2006).
64. Touzet, H. & Varre, J. S. Efficient and accurate P-value computation for Position Weight Matrices. *Algorithms Mol. Biol.* **2**, 15 (2007).
65. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).

Crystal structure of a phosphorylation-coupled saccharide transporter

Yu Cao¹, Xiangshu Jin^{2*}, Elena J. Levin^{1*}, Hua Huang^{1*}, Yinong Zong^{3*}, Matthias Quick^{4,5}, Jun Weng¹, Yaping Pan¹, James Love⁶, Marco Punta^{6,7}, Burkhard Rost^{6,7}, Wayne A. Hendrickson^{6,8}, Jonathan A. Javitch^{4,5,9}, Kanagalaghatta R. Rajashankar¹⁰ & Ming Zhou¹

Saccharides have a central role in the nutrition of all living organisms. Whereas several saccharide uptake systems are shared between the different phylogenetic kingdoms, the phosphoenolpyruvate-dependent phosphotransferase system exists almost exclusively in bacteria. This multi-component system includes an integral membrane protein EIIC that transports saccharides and assists in their phosphorylation. Here we present the crystal structure of an EIIC from *Bacillus cereus* that transports diacetylchitobiose. The EIIC is a homodimer, with an expansive interface formed between the amino-terminal halves of the two protomers. The carboxy-terminal half of each protomer has a large binding pocket that contains a diacetylchitobiose, which is occluded from both sides of the membrane with its site of phosphorylation near the conserved His 250 and Glu 334 residues. The structure shows the architecture of this important class of transporters, identifies the determinants of substrate binding and phosphorylation, and provides a framework for understanding the mechanism of sugar translocation.

Bacterial phosphoenolpyruvate-dependent phosphotransferase systems (PTSs)¹ transport saccharides across the cell membrane and phosphorylate them before their release into the cytosol^{2–7}. Phosphorylation of the incoming saccharide primes it for subsequent utilization as a nutrient in cellular metabolism and also prevents its efflux. Although the system can transport a cognate sugar by slow facilitated diffusion *in vitro* in the absence of phosphorylation, phosphorylation greatly speeds up the overall rate of sugar uptake^{2,3}, allows concentration of intracellular substrate relative to the environment, and is necessary for growth of the host bacteria when the PTS sugar is provided as the sole carbon source. Unlike the primary ABC-type transporters that hydrolyse ATP^{8,9} or the secondary transporters that harness a sodium or proton gradient^{10–12} to drive transport, PTSs therefore use covalent modification of their substrate during transport to ensure its unidirectional flux.

PTSs are composed of three components: enzyme I (EI), the heat-stable phosphocarrier protein (HPr) and enzyme II (EII). Both EI and HPr are general energy-coupling proteins and are not sugar specific, whereas EII is sugar specific and is itself a protein complex composed of the cytosolic EIIA and EIIB proteins and the integral membrane protein EIIC. In certain EIIs, EIIA or EIIB or both are translated with EIIC as a single polypeptide chain. Bacteria often possess several different types of EIIs that are induced by the presence of their substrate^{13,14}. The phosphate group originates from phosphoenolpyruvate, and is transferred sequentially to EI, HPr, EIIA, EIIB and eventually to the incoming sugar substrate bound to EIIC, the component responsible for translocating the sugar^{2,3}.

Of the four EIIC superfamilies, the largest is the Glc family, which has subfamilies each specialized in transporting glucose, several β -glucosides, mannitol, fructose, or lactose^{6,15}. All Glc family EIICs have an almost universally conserved glutamate residue (Supplementary

Fig. 1) essential for substrate transport and phosphorylation^{16,17}. This conserved glutamate is located within a conserved motif, which was first identified as GITEP in the glucose- and β -glucoside-specific EIICs^{2,6}. To understand further the mechanism of sugar selectivity, translocation and phosphorylation, we initiated structural studies on a group of EIICs that are members of the lactose subfamily of the Glc superfamily. These members have an orthologue in *Escherichia coli*, ChbC, which was shown to transport *N,N'*-diacetylchitobiose ((GlcNAc)₂), a β -1,4-linked *N*-acetylglucosamine disaccharide^{18,19}. We crystallized and solved the structure of a ChbC homologue from *Bacillus cereus* (Supplementary Fig. 1).

Functional characterization

The *B. cereus* *chbC* gene was heterologously expressed in *E. coli* and the resulting protein purified to homogeneity (Supplementary Fig. 2a). The molecular weight of detergent-solubilized ChbC was estimated at approximately 100 kDa by combining size-exclusion chromatography with light scattering and refractive index measurements. As the molecular weight of an individual protomer is ~47 kDa, detergent-solubilized ChbC is therefore a homodimer. This dimer is stable and monodispersed in a number of detergents, but partially dissociates in shorter-chain detergents such as octylmaltoside (Supplementary Fig. 2b–e). These results are consistent with earlier reports of a dimeric assembly for EIICs from other members of the Glc superfamily^{20–24} and indicate that the purified ChbC retains its native quaternary assembly in long-chain detergents.

We reconstituted purified *B. cereus* ChbC into proteoliposomes and measured its ability to transport sugars by monitoring uptake of ¹⁴C-labelled *N*-acetylglucosamine (GlcNAc), which is the monosaccharide that is condensed to form (GlcNAc)₂. Addition of ¹⁴C-GlcNAc

¹Department of Physiology & Cellular Biophysics, College of Physicians and Surgeons, Columbia University, 630 West 168th Street, New York, New York 10032, USA. ²Center for Computational Biology and Bioinformatics, Department of Biochemistry and Molecular Biophysics, Howard Hughes Medical Institute, Columbia University, 1130 St. Nicholas Ave, Room 815, New York, New York 10032, USA.

³Sanford-Burnham Institute, La Jolla, California 92037, USA. ⁴Department of Psychiatry and Center for Molecular Recognition, Columbia University, 630 West 168th Street, New York, New York 10032, USA. ⁵New York State Psychiatric Institute, Division of Molecular Therapeutics, 1051 Riverside Drive, New York, New York 10032, USA. ⁶New York Consortium on Membrane Protein Structure, New York Structural Biology Center, 89 Convent Avenue, New York, New York 10027, USA. ⁷Department of Computer Science and Institute for Advanced Study, Technical University of Munich, D-85748 Munich, Germany. ⁸Department of Biochemistry and Molecular Biophysics, Howard Hughes Medical Institute, Columbia University, 630 West 168th Street, New York, New York 10032, USA. ⁹Department of Pharmacology, Columbia University, 630 West 168th Street, New York, New York 10032, USA. ¹⁰Department of Chemistry and Chemical Biology, Cornell University, NE-CAT, Advanced Photon Source, Argonne, Illinois 60439, USA.

*These authors contributed equally to this work.

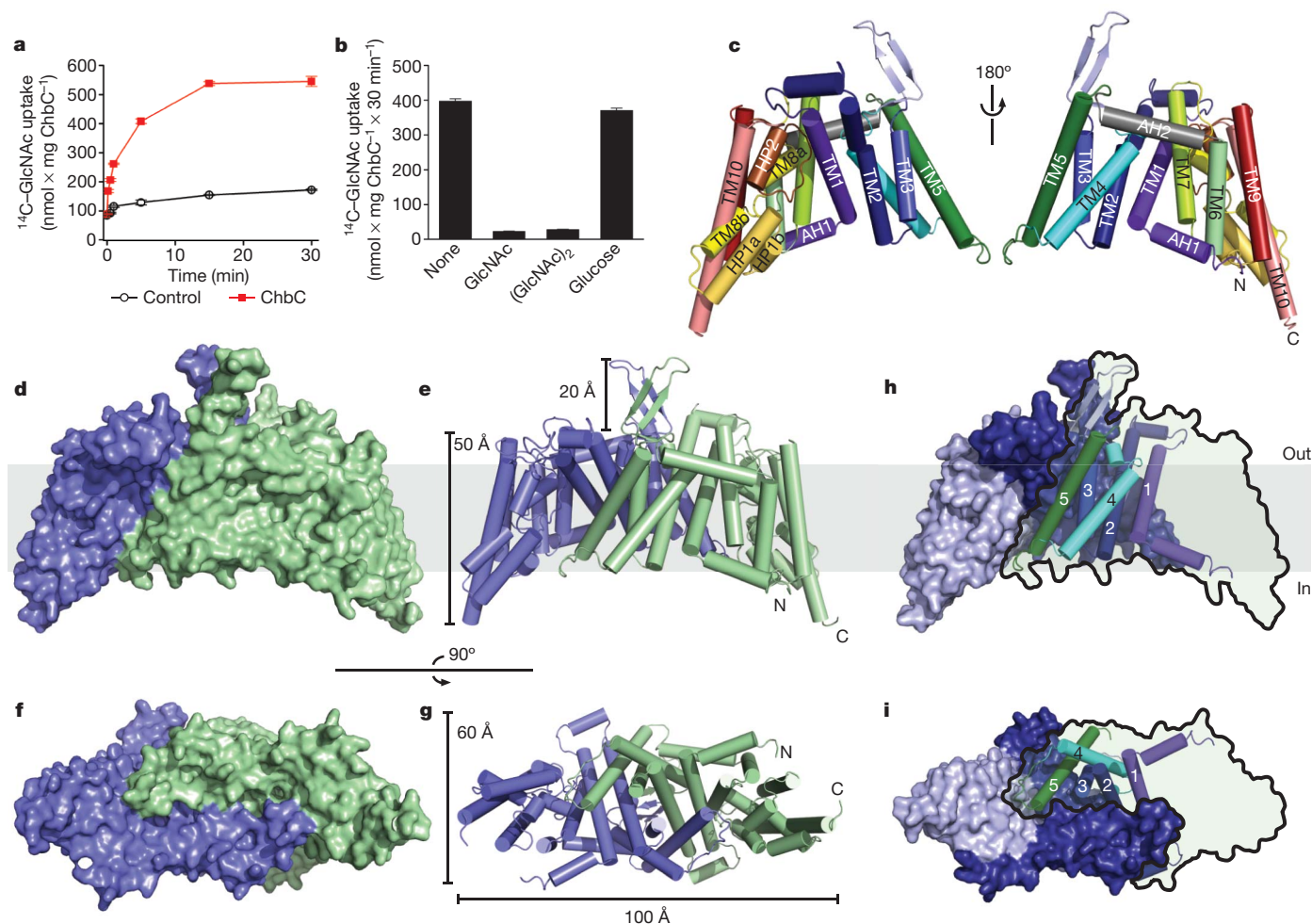


Figure 1 | Function and structure of ChbC. **a**, Time course of the uptake of $92 \mu\text{M}$ ^{14}C -labelled GlcNAc in proteoliposomes reconstituted with *B. cereus* ChbC (red squares) or in control liposomes (black circles). **b**, Accumulation after 30 min of ^{14}C -labelled GlcNAc in ChbC-containing proteoliposomes in the absence or presence of 10 mM unlabelled GlcNAc, (GlcNAc) $_2$, or glucose. Error bars in **a** and **b** are s.e.m. of 3–6 measurements. **c**, A ChbC protomer is shown from two orientations. **d–g**, The structure of the ChbC dimer is shown in

cartoon and surface representations as viewed from within the plane of the membrane, represented as a grey rectangle (**d**, **e**), or the intracellular side of the membrane (**f**, **g**). **h**, **i**, The same views of the ChbC dimer, but with the back protomer shown as an opaque surface and the front protomer as an outline. Helices TM1–5 are shown as either a darker surface (back) or as coloured cylinders (front).

resulted in a time-dependent accumulation of the radiotracer within the lumen of the proteoliposomes that stabilized after ~ 15 min (Fig. 1a), consistent with a facilitated diffusion process that dissipates the initial concentration gradient of the radiotracer. Control liposomes lacking ChbC showed little accumulation even after 30 min. Further experiments showed that ^{14}C -GlcNAc uptake was significantly inhibited in the presence of non-labelled GlcNAc or (GlcNAc) $_2$, whereas the same concentration of glucose had no significant effect (Fig. 1b). This experiment indicates that the reconstituted ChbC is capable of translocating a sugar, and that it is selective for GlcNAc and (GlcNAc) $_2$ over glucose.

Structure determination

After extensive refinement of crystallization conditions, a data set was collected on a crystal grown in the presence of 4 mM (GlcNAc) $_2$. The crystal had $P4_32_12$ symmetry and diffracted to 3.3 Å (Supplementary Table 1). Initial phases were estimated from a Ta $_6$ Br $_{12}$ derivative diffracting to 4.5 Å, and the phases were gradually extended to the native data set (Methods and Supplementary Fig. 3). There are four ChbC protomers in the asymmetric unit, and the building and refinement of an accurate atomic model were aided by the use of four-fold non-crystallographic symmetry (NCS) restraints, which were

maintained throughout the refinement until the last few rounds (Methods). The final models of all four protomers in the asymmetric unit contain the full-length ChbC except for two residues at the N terminus that are not resolved. In addition, each chain contains one (GlcNAc) $_2$ and two nonylmaltosides (NM) that were used to solubilize ChbC. One of the NM molecules is only partially resolved.

Tertiary and quaternary structure of ChbC

Each ChbC protomer contains 10 transmembrane helical regions (TM1–10), including one (TM8) that is tilted at a roughly 45° angle to the membrane norm and is split into two short hydrophobic helices joined by a hydrophilic loop (Fig. 1c and Supplementary Fig. 4). It also contains two re-entrant hairpin-like structures (HP1 and HP2) with opposite orientations in the membrane, and two horizontal amphipathic α -helices (AH1 and AH2). AH1 and AH2 probably lie along the inner and outer boundaries of the hydrophobic core of the lipid bilayer, which is marked in Fig. 1d and e. To the best of our knowledge, ChbC has a novel fold.

The protein is a homodimer and the two protomers are oriented parallel in the membrane, related by a two-fold axis perpendicular to the membrane (Fig. 1d, e and Supplementary Fig. 5). Both the N and C termini probably reside on the cytoplasmic side (Supplementary Fig. 4)

as inferred from the ‘positive-inside’ rule²⁵, and from the experimentally determined topology of ChbC from *E. coli*²⁶. This assignment is also consistent with the location of the termini determined experimentally in other members of the Glc superfamily^{6,27–29}. When viewed from within the membrane with the extracellular side on the top, the dimer is roughly 50 Å thick along the two-fold axis and has the shape of a capsized canoe, with a concave surface facing the intracellular side (Fig. 1d, e). β -Hairpins from each protomer protrude an extra 20 Å into the extracellular space, although the hairpins mediate a key crystal contact and may be perturbed slightly from their native conformation (Supplementary Fig. 6). When the dimer is viewed looking down the two-fold axis from the intracellular side, the two dimensions of the concave surface are ~60 Å and 100 Å (Fig. 1f, g). Stereo views of the ChbC dimer in three orientations are shown in Supplementary Fig. 5.

The extensive dimer interface is formed primarily by the N-terminal half of ChbC: TM1, 2, 3 and 5 from each protomer line the interface with a buried surface area of 2,746 Å² per protomer (Fig. 1h and Supplementary Fig. 7). The long loop between TM4 and 5 also contributes to the interface by extending along the cytoplasmic face of the neighbouring subunit (Fig. 1i). This large and mostly hydrophobic interface is expected because EIICs are known to function as dimers in the membrane^{20–24}. An extensive dimer interface was also observed in an electron microscopy projection map of a Glc family EIIC that transports mannitol³⁰, indicating that this feature is conserved among subfamilies of the Glc family transporters.

Substrate-binding site

The C-terminal half of each protomer (TM6–10) contains a deep, electronegative cleft on its intracellular side (Fig. 2a). Although the cleft is located on the intracellular face of each protomer, it is not solvent-exposed because part of TM5 and the preceding TM4–5 loop from the neighbouring protomer extend beneath it (Fig. 1i). This cleft is lined partly by the re-entrant hairpin loops HP1 and HP2. HP1 harbours the glutamate residue (Glu 334) in the Glc family conserved motif, which is NINEP in this ChbC (Supplementary Fig. 1). The tips of HP1 and HP2 meet in the middle of the membrane (Fig. 2b and Supplementary Fig. 8a). The arrangement of these two loops is strongly reminiscent of two re-entrant hairpins in the otherwise dissimilar structure of the glutamate transporter Glt_{ph}³¹.

A large non-protein electron density is present in the deep cleft (Fig. 2b), and although we cannot unambiguously determine its identity owing to the modest resolution of the data set, this electron density is consistent with the size and shape of a (GlcNAc)₂ molecule. The shape of the ligand density allows for two orientations of (GlcNAc)₂, with the non-reducing sugar either closer to or farther away from the

intracellular side (Supplementary Fig. 9). It is known that *E. coli* ChbC phosphorylates (GlcNAc)₂ on the sixth position hydroxyl of the non-reducing sugar (C6-OH)¹⁹. If (GlcNAc)₂ were oriented with its non-reducing sugar ring closer to the intracellular side (Fig. 2c and Supplementary Fig. 9a), the C6-OH would be accessible for phosphorylation. This orientation would also place the C6-OH within hydrogen-bonding distance of the conserved residues Glu 334 and His 250, whose importance for sugar binding and phosphorylation has been demonstrated in an EIIC that transports mannitol^{16,17,32–34}. In contrast, the alternative orientation would position the C6-OH in the protein interior (Supplementary Fig. 9b) where it would not seem accessible for the required phosphorylation. In light of these observations, we deemed the former orientation to be more plausible and used it in the final model. After refinement, the hydroxyl oxygen of C6-OH is 2.6–2.8 Å from the carboxylate of Glu 334 on HP1, and 2.7–3.1 Å from the ϵ -nitrogen on His 250, which is part of the loop between TM6–7 (Fig. 2c and Supplementary Fig. 8b). We speculate that Glu 334, His 250 and the C6-OH may form part of an active site where transfer of a phosphate group from EIIB takes place, although the precise mechanism of catalysis is currently unknown.

Although the substrate selectivity of ChbC has not been measured systematically, it appears that the binding pocket is well suited for (GlcNAc)₂. In addition to Glu 334 and His 250, the side chains of conserved Trp 245 from TM7, Asp 290 from TM8a and Asn 333 from HP1 are able to form hydrogen bonds with the (GlcNAc)₂ when it is modelled in the orientation shown in Fig. 2c. Trp 382 from HP2 provides stacking interactions with the ring of the reducing sugar. In addition, the acetamide group on the non-reducing sugar makes two interactions with the protein: the backbone carbonyl oxygen atom of Gly 297 from the conserved TM8 loop makes a hydrogen bond with the nitrogen atom, and the aromatic ring of Tyr 294, which is also from the TM8 loop, is ~3.5 Å from the methyl group. When a glucose is modelled in the binding site by aligning it with the non-reducing sugar of (GlcNAc)₂, the interactions between the acetamide group on (GlcNAc)₂ and residues Trp 245 and Tyr 294 are both missing, suggesting that these two interactions are important for sugar selectivity (Fig. 1b). Curiously, the substrate-binding cavity is substantially larger than necessary to accommodate a (GlcNAc)₂ molecule (Fig. 2a). In *E. coli*, ChbC was shown to also transport the trisaccharide of GlcNAc¹⁹, and the large size of this cavity indicates that *B. cereus* ChbC is able to accommodate a trisaccharide as well.

Implications for mechanism of transport

In the observed conformation, the binding pocket for (GlcNAc)₂ faces the cytoplasmic side, but the bound (GlcNAc)₂ cannot diffuse to either

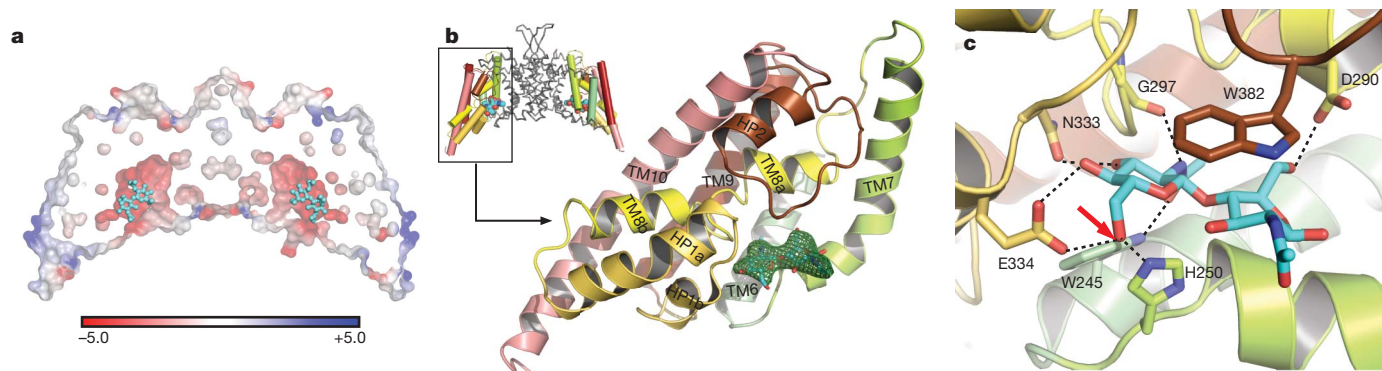


Figure 2 | The C-terminal sugar-binding domain. **a**, Cross-section of the solvent-accessible surface of the ChbC dimer, coloured by electrostatics as calculated by the program DelPhi⁴¹. Bound (GlcNAc)₂ molecules are shown in cyan. **b**, The C-terminal domain and bound (GlcNAc)₂ molecule viewed from the plane of the membrane. The green mesh corresponds to $F_o - F_c$ density calculated in the absence of (GlcNAc)₂ and contoured at 2.5σ . The inset on the

upper left shows the location of the C-terminal domain in the dimer. **c**, The sugar-binding pocket viewed from the intracellular side. A (GlcNAc)₂ molecule is shown modelled in the orientation placing the C6-OH of the non-reducing sugar (red arrow) closest to the cytoplasm, along with residues potentially forming hydrogen bonds or hydrophobic interactions with the sugar.

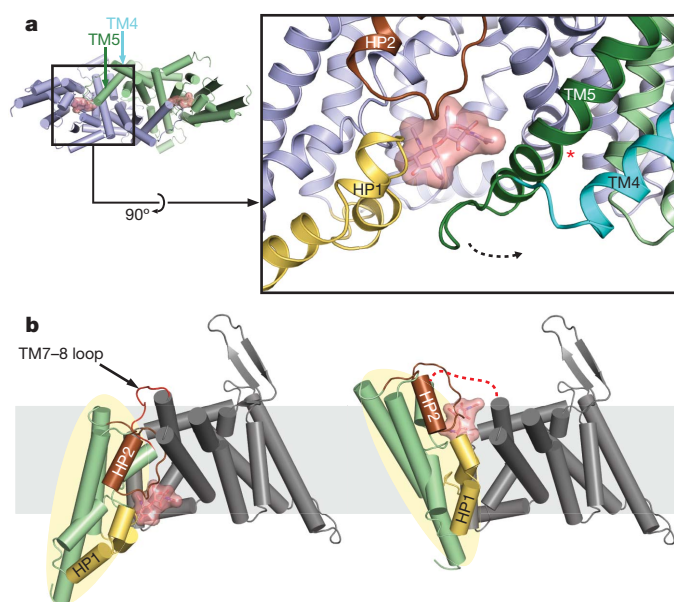


Figure 3 | Proposed conformational changes in sugar transport. **a**, On the intracellular face of ChbC (left panel), helix TM5 and the TM4–5 loop from one protomer cover the binding pocket for (GlcNAc)₂ (shown as a pink surface) on the opposite protomer. The right panel corresponds to the region marked on the left with a rectangle, zoomed in and rotated to view from within the plane of the membrane. Straightening of a kink in TM5 (red star) could potentially expose the substrate-binding site to the cytoplasm. **b**, The substrate is occluded (left panel, crystal structure) from the periplasmic side by the highlighted region containing TM8–10 (green), HP1 (orange) and HP2 (brown), connected to the remainder of the protein by the TM7–8 loop (red). A rigid-body rotation of this region could potentially move and expose the substrate-binding site to the periplasmic space (right panel, model). The short helix between TM1 and TM2 is omitted for clarity.

side of the membrane without changes in protein conformation. The crystal structure therefore probably corresponds to what is referred to as the occluded state in the terminology of the alternating access model proposed for sodium-coupled secondary transporters^{35–38}. In this framework, the full transport cycle should include at minimum two additional states: an outward-open state capable of binding substrate from the periplasm, and an inward-open state that interacts with EIIB to phosphorylate and release the substrate into the cytoplasm. On the basis of the known structure, we will briefly speculate on possible conformational changes leading to the other states.

The substrate-binding cavity is sealed off to the intracellular side by residues in the loop between TM4 and 5 from the neighbouring protomer (Fig. 3a). The TM4–5 loop has little interaction with the rest of the protein and could potentially be moved away to expose the bound substrate by straightening a kink near the N terminus of TM5 (Fig. 3a and Supplementary Fig. 10). Therefore, the TM4–5 loop seems a reasonable candidate for the intracellular gate. Once the substrate is released the strong electronegativity of the binding cavity (Fig. 2a) may assist in preventing the phosphorylated sugar from rebinding to the transporter and effluxing from the cell. Although it is impossible to determine from the crystal structure alone, the involvement of structural features from both protomers in forming the binding site, along with the considerable size of the dimer interface, raises the interesting possibility that binding or release of substrate may be cooperative. Further functional and structural studies, and in particular the structure of ChbC in complex with its corresponding EIIB ChbB in the phosphorylated state, will be necessary to reveal the nature and sequence of the conformational changes leading to phosphorylation and release of the bound carbohydrate.

Because the substrate-binding site is located nearer to the cytoplasmic side of the membrane, it would require a more substantial

conformational change to form the outward-open state. However, similarities between ChbC and the unrelated transporter Glt_{ph} provide one possible clue. In that transporter, a rigid-body motion of a transport domain containing the substrate-binding pocket relative to an immobile oligomerization domain is responsible for conversion between the outward and inward-facing states. The oppositely oriented re-entrant loops on the transport domain form both the substrate-binding site and the moving interface with the oligomerization domain³⁷. The parallels between the architecture of the transport domain in Glt_{ph} and the C-terminal region of ChbC containing HP1, HP2 and TM8–10 raise the intriguing possibility that a similar rigid-body motion in ChbC could be responsible for converting the inward-occluded state observed in the crystal structure to an outward-open state with an exposed binding site for external sugar (Fig. 3b and Supplementary Fig. 10). This large motion would be facilitated by the generous length of the 12-residue extracellular loop between TM7 and TM8. Further structural studies will be necessary to resolve the nature of the outward-facing state, possibly with apo-ChbC, which probably favours a conformation capable of binding periplasmic substrate.

METHODS SUMMARY

B. cereus ChbC was cloned into a modified pET plasmid (Novagen) with a C-terminal polyhistidine tag connected by a TEV protease recognition site³⁹. The *chbC* gene was overexpressed in BL21(DE3) cells and the protein purified on an IMAC column. After cleavage of the tag by TEV protease, the protein was then exchanged into buffer containing 150 mM NaCl, 20 mM HEPES, pH 7.5, 5 mM β-mercaptoethanol and 12 mM *n*-nonyl-β-D-maltoside, and concentrated to 6 mg ml^{−1}. Crystals in the P4₃2₁2 space group were grown by the sitting-drop method in solution containing 4 mM *N,N'*-diacetylchitobiose, 30% polyethyleneglycol (PEG) 400, 100 mM Li₂SO₄, 0.5% polyvinylpyrrolidone, and 100 mM sodium citrate, pH 5.6. Diffraction data were collected and phased by the single-wavelength anomalous dispersion (SAD) method using Ta₆Br₁₂-derivatized P4₃2₁2 crystals. Experimental phases were obtained to 4.5 Å and improved by solvent flattening and averaging, and iterative rounds of model building and refinement were then carried out to obtain the final model.

For the functional assays, ChbC was reconstituted in liposomes following a method described for a K⁺ channel⁴⁰. Uptake of ¹⁴C-GlcNAc (45 Ci mmol^{−1}; Moravsek) was measured in buffer containing 100 mM potassium phosphate, pH 7.5 for varying periods of time. The reactions were quenched with ice-cold buffer containing 100 mM potassium phosphate, pH 6.0 and 100 mM LiCl, and immediately filtered through GF/F filters (Advantec MFS). GlcNAc uptake was quantified by comparing scintillation counts of the filters with standard curves from known amounts of ¹⁴C-GlcNAc.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 September 2010; accepted 11 February 2011.

Published online 6 April 2011.

- Kundig, W., Ghosh, S. & Roseman, S. Phosphate bound to histidine in a protein as an intermediate in a novel phospho-transferase system. *Proc. Natl Acad. Sci. USA* **52**, 1067–1074 (1964).
- Postma, P. W., Lengeler, J. W. & Jacobson, G. R. Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. *Microbiol. Rev.* **57**, 543–594 (1993).
- Postma, P. W. & Lengeler, J. W. Phosphoenolpyruvate:carbohydrate phosphotransferase system of bacteria. *Microbiol. Rev.* **49**, 232–269 (1985).
- Siebold, C., Flückiger, K., Beutler, R. & Erni, B. Carbohydrate transporters of the bacterial phosphoenolpyruvate: sugar phosphotransferase system (PTS). *FEBS Lett.* **504**, 104–111 (2001).
- Robillard, G. T. & Broos, J. Structure/function studies on the bacterial carbohydrate transporters, enzymes II, of the phosphoenolpyruvate-dependent phosphotransferase system. *Biochim. Biophys. Acta* **1422**, 73–104 (1999).
- Nguyen, T. X., Yen, M. R., Barabote, R. D. & Saier, M. H. Jr. Topological predictions for integral membrane permeases of the phosphoenolpyruvate:sugar phosphotransferase system. *J. Mol. Microbiol. Biotechnol.* **11**, 345–360 (2006).
- Meadow, N. D., Fox, D. K. & Roseman, S. The bacterial phosphoenolpyruvate: glycolate phosphotransferase system. *Annu. Rev. Biochem.* **59**, 497–542 (1990).
- Rees, D. C., Johnson, E. & Lewinson, O. ABC transporters: the power to change. *Nature Rev. Mol. Cell Biol.* **10**, 218–227 (2009).
- Davidson, A. L., Dassa, E., Orelle, C. & Chen, J. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol. Mol. Biol. Rev.* **72**, 317–364 (2008).

10. Krishnamurthy, H., Piscitelli, C. L. & Gouaux, E. Unlocking the molecular secrets of sodium-coupled transporters. *Nature* **459**, 347–355 (2009).
11. Lemieux, M. J., Huang, Y. & Wang, D. N. The structural basis of substrate translocation by the *Escherichia coli* glycerol-3-phosphate transporter: a member of the major facilitator superfamily. *Curr. Opin. Struct. Biol.* **14**, 405–412 (2004).
12. Abramson, J., Kaback, H. R. & Iwata, S. Structural comparison of lactose permease and the glycerol-3-phosphate antiporter: members of the major facilitator superfamily. *Curr. Opin. Struct. Biol.* **14**, 413–419 (2004).
13. Rephaeli, A. W. & Saier, M. H. Jr. Regulation of genes coding for enzyme constituents of the bacterial phosphotransferase system. *J. Bacteriol.* **141**, 658–663 (1980).
14. Deutscher, J., Francke, C. & Postma, P. W. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol. Mol. Biol. Rev.* **70**, 939–1031 (2006).
15. Saier, M. H., Hvorup, R. N. & Barabote, R. D. Evolution of the bacterial phosphotransferase system: from carriers and enzymes to group translocators. *Biochem. Soc. Trans.* **33**, 220–224 (2005).
16. Saraceni-Richards, C. A. & Jacobson, G. R. Subunit and amino acid interactions in the *Escherichia coli* mannitol permease: a functional complementation study of coexpressed mutant permease proteins. *J. Bacteriol.* **179**, 5171–5177 (1997).
17. Saraceni-Richards, C. A. & Jacobson, G. R. A conserved glutamate residue, Glu-257, is important for substrate binding and transport by the *Escherichia coli* mannitol permease. *J. Bacteriol.* **179**, 1135–1142 (1997).
18. Keyhani, N. O. & Roseman, S. Wild-type *Escherichia coli* grows on the chitin disaccharide, *N,N'*-diacetylchitobiose, by expressing the *cel* operon. *Proc. Natl Acad. Sci. USA* **94**, 14367–14371 (1997).
19. Keyhani, N. O., Wang, L. X., Lee, Y. C. & Roseman, S. The chitin disaccharide, *N,N'*-diacetylchitobiose, is catabolized by *Escherichia coli* and is transported/ phosphorylated by the phosphoenolpyruvate:glycose phosphotransferase system. *J. Biol. Chem.* **275**, 33084–33090 (2000).
20. van Montfort, B. A. *et al.* Mapping of the dimer interface of the *Escherichia coli* mannitol permease by cysteine cross-linking. *J. Biol. Chem.* **277**, 14717–14723 (2002).
21. Erni, B. Glucose-specific permease of the bacterial phosphotransferase system: phosphorylation and oligomeric structure of the glucose-specific II^{Glc}-III^{Glc} complex of *Salmonella typhimurium*. *Biochemistry* **25**, 305–312 (1986).
22. Pas, H. H., Ellory, J. C. & Robillard, G. T. Bacterial phosphoenolpyruvate-dependent phosphotransferase system: association state of membrane-bound mannitol-specific enzyme II demonstrated by inactivation. *Biochemistry* **26**, 6689–6696 (1987).
23. Khandekar, S. S. & Jacobson, G. R. Evidence for two distinct conformations of the *Escherichia coli* mannitol permease that are important for its transport and phosphorylation functions. *J. Cell. Biochem.* **39**, 207–216 (1989).
24. Chen, Q. & Amster-Choder, O. BglF, the sensor of the *bgl* system and the β -glucosides permease of *Escherichia coli*: evidence for dimerization and intersubunit phosphotransfer. *Biochemistry* **37**, 8714–8723 (1998).
25. von Heijne, G. & Gavel, Y. Topogenic signals in integral membrane proteins. *Eur. J. Biochem.* **174**, 671–678 (1988).
26. Daley, D. O. *et al.* Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science* **308**, 1321–1323 (2005).
27. Sugiyama, J. E., Mahmoodian, S. & Jacobson, G. R. Membrane topology analysis of *Escherichia coli* mannitol permease by using a nested-deletion method to create *mtlA-phoA* fusions. *Proc. Natl Acad. Sci. USA* **88**, 9603–9607 (1991).
28. Buhr, A. & Erni, B. Membrane topology of the glucose transporter of *Escherichia coli*. *J. Biol. Chem.* **268**, 11599–11603 (1993).
29. Yagur-Kroll, S. & Amster-Choder, O. Dynamic membrane topology of the *Escherichia coli* β -glucoside transporter BglF. *J. Biol. Chem.* **280**, 19306–19318 (2005).
30. Koning, R. I. *et al.* The 5 Å projection structure of the transmembrane domain of the mannitol transporter enzyme II. *J. Mol. Biol.* **287**, 845–851 (1999).
31. Yernool, D., Boudker, O., Jin, Y. & Gouaux, E. Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*. *Nature* **431**, 811–818 (2004).
32. Weng, Q. P. & Jacobson, G. R. Role of a conserved histidine residue, His-195, in the activities of the *Escherichia coli* mannitol permease. *Biochemistry* **32**, 11211–11216 (1993).
33. Weng, Q. P., Elder, J. & Jacobson, G. R. Site-specific mutagenesis of residues in the *Escherichia coli* mannitol permease that have been suggested to be important for its phosphorylation and chemoreception functions. *J. Biol. Chem.* **267**, 19529–19535 (1992).
34. Opacić, M., Vos, E. P., Hesp, B. H. & Broos, J. Localization of the substrate-binding site in the homodimeric mannitol transporter, EII^{mtl}, of *Escherichia coli*. *J. Biol. Chem.* **285**, 25324–25331 (2010).
35. Widdas, W. F. Inability of diffusion to account for placental glucose transfer in the sheep and consideration of the kinetics of a possible carrier transfer. *J. Physiol. (Lond.)* **118**, 23–39 (1952).
36. Yamashita, A. *et al.* Crystal structure of a bacterial homologue of Na⁺/Cl[−]-dependent neurotransmitter transporters. *Nature* **437**, 215–223 (2005).
37. Reyes, N., Ginter, C. & Boudker, O. Transport mechanism of a bacterial homologue of glutamate transporters. *Nature* **462**, 880–885 (2009).
38. Shi, L. *et al.* The mechanism of a neurotransmitter:sodium symporter—inward release of Na⁺ and substrate is triggered by substrate in a second binding site. *Mol. Cell* **30**, 667–677 (2008).
39. Love, J. *et al.* The New York Consortium on Membrane Protein Structure (NYCOMPS): a high-throughput platform for structural genomics of integral membrane proteins. *J. Struct. Funct. Genomics* **11**, 191–199 (2010).
40. Zhou, M. & MacKinnon, R. A mutant KcsA K⁺ channel with altered conduction properties and selectivity filter ion distribution. *J. Mol. Biol.* **338**, 839–846 (2004).
41. Rocchia, W. *et al.* Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J. Comput. Chem.* **23**, 128–137 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Data for this study were measured at beamlines X4A, X4C, X25 and X29 of the National Synchrotron Light Source and the NE-CAT 24ID-C and E at the Advanced Photon Source. This work was supported by the US National Institutes of Health (DK088057, GM098878 and GM05026-sub0007 to M.Z., and T32HL087745 to E.J.L.). M.Z. is a Pew Scholar in Biomedical Sciences. The NYCOMPS central facility was supported by GM05026 to W.A.H. as part of the Protein Structure Initiative (PSI-2) established by the National Institute of General Medical Sciences. The authors would like to thank B. Honig for support and M. Saier, B. Erni, R. Kaback and D.-N. Wang for comments on the manuscript and helpful discussions. M.Z. is grateful to R. MacKinnon for advice and encouragement.

Author Contributions J.L., M.P., B.R. and W.A.H. identified ChbC homologues in the database. J.L. carried out the cloning and the initial expression studies. Y.C., H.H., E.J.L., J.W. and M.Z. performed protein expression, purification, crystallization and X-ray diffraction data collection and analysis. X.J., Y.Z., E.J.L. and M.Z. solved and refined the structures. M.Q., Y.C., Y.P., J.A.J. and M.Z. characterized ChbC function. K.R.R. advised on data collection and crystallography. E.J.L. and M.Z. wrote the manuscript with input from all authors.

Author Information Atomic coordinates and structure factors have been deposited with the Protein Data Bank under accession code 3QNQ. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.Z. (mz2140@columbia.edu).

METHODS

Target selection, cloning and initial protein production. ChbC was established as a pipeline target for structural studies by a bioinformatics analysis^{39,42}. A total of 25 *chbC* genes from 13 prokaryotic genomes were identified, amplified by PCR from the genomic DNA, and cloned into a modified pET plasmid (Novagen) with a C-terminal deca-histidine tag and a TEV protease recognition site. *chbC* genes were then overexpressed in *E. coli* BL21(DE3) cells in small-scale cultures (~1 ml), and their translation level examined using western blots. The target selection, cloning, and protein production screening were performed at the central facility of the New York Consortium on Membrane Protein Structure (NYCOMPS) as described previously³⁹.

Protein purification and crystallization. Seven western-positive clones received from the NYCOMPS were scaled up for mid-scale (1 l) purification studies. ChbCs from *Salmonella enterica* and *B. cereus* yielded higher than 0.25 mg l⁻¹ cell culture. Although both proteins exhibited a monodispersed profile in size-exclusion chromatography, only ChbC from *B. cereus* (ChbC) produced diffracting crystals and thus became the focus of crystallization efforts. After cleavage of the deca-histidine tag, the protein contains the full-length ChbC protein, residues 1 to 434, plus nine residues (AAAENLYFQ) at the C terminus due to addition of a cloning site and the TEV protease recognition site.

For large-scale (10–20 l) production and purification of native ChbC, cells were grown in Luria broth at 37 °C and induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) after the OD_{600 nm} reached ~1.0. ChbC extraction and purification followed a protocol described in ref. 43. After removal of the His tag with TEV protease, the protein was concentrated to ~6 mg ml⁻¹ and subjected to size exclusion chromatography with a Superdex 200 10/300 GL column (GE Health Sciences) equilibrated in 150 mM NaCl, 20 mM HEPES, pH 7.5, 5 mM β-mercaptoethanol and 12 mM *n*-nonyl-β-D-maltoside (NM). Purified ChbC protein was concentrated to ~10 mg ml⁻¹ as approximated by ultraviolet absorbance.

ChbC crystals with P₄₃2₁ symmetry were grown over a period of two weeks or longer by vapour diffusion in sitting drops mixed from 2–3 μl of the protein solution supplemented with 4 mM *N,N'*-diacetylchitobiose and an equal volume of well solution containing 30% polyethyleneglycol (PEG) 400, 100 mM Li₂SO₄, 0.5% polyvinylpyrrolidone and 100 mM sodium citrate, pH 5.6. Tantalum-derivatized crystals were prepared by adding Ta₂Br₁₂ powder into sitting drops containing ChbC crystals. The crystals gradually turned green after 24–48 h, and were directly flash-frozen in liquid nitrogen for X-ray diffraction. The P₄₃2₁ crystals diffracted to resolutions of up to 3.3 Å and the Tantalum-derivatized crystals to 4.5 Å.

Transport measurements in proteoliposomes. Purified ChbC was reconstituted at a 1:100 (w/w) ratio into liposomes composed of 1-palmitoyl-2-oleoyl-phosphatidylethanolamine and 1-palmitoyl-2-oleoyl-phosphatidylglycerol (Avanti Polar Lipids) in a ratio of 3:1 (w/w) in 100 mM potassium phosphate, pH 7.5 as described previously⁴⁰. Before the uptake reaction, frozen ChbC-containing proteoliposomes and control liposomes (at 10 mg lipid per ml) were subjected to three freeze/thaw cycles followed by extrusion through a 400 nm polycarbonate membrane (Avestin). Uptake of 92 μM ¹⁴C-GlcNAc (45 Ci mmol⁻¹; Moravsek) was measured at 23 °C in assay buffer composed of 100 mM potassium phosphate buffer, pH 7.5 for the indicated periods of time. The reactions were stopped by quenching the samples with ice-cold 100 mM potassium phosphate, pH 6.0/100 mM LiCl, followed by rapid filtration through GF/F filters (Advantec MFS) and scintillation

counting of the filters. Known amounts of ¹⁴C-GlcNAc were used to convert the amount of internalized radioactivity into nmol per mg of protein. The amount of ChbC in the proteoliposomes used for the uptake reaction was determined⁴⁴.

Data collection and structure solution. Diffraction data were collected on beamlines X25 and X29 at the National Synchrotron Light Source and 24ID-C and 24ID-E at the Advanced Photon Source. Owing to the long *c*-axis, crystals were re-oriented using mini-kappa to avoid overlaps. The data were indexed, integrated and scaled using HKL2000⁴⁵. The experimental phases were determined to 4.5 Å by SAD using a data set collected at the tantalum L-III absorption edge on a crystal derivatized with Ta₆Br₁₂. The positions of 12 Ta sites corresponding to two clusters were located using SHELXD⁴⁶ and refined with SHARP⁴⁷. The experimental phases were calculated using SHARP and improved by solvent flattening with DM⁴⁸. The resultant density-modified map allowed for identification of molecular boundaries for four chains in the asymmetric unit by manual inspection. Within the boundaries of each of the four chains in the asymmetric unit, Cα traces for several helical fragments were positioned manually at equivalent regions. The Cα coordinates of these fragments were then used to calculate non-crystallographic symmetry (NCS) operators. Subsequent four-fold NCS averaging, solvent flattening and histogram matching were done with DM to extend phases to a 3.3 Å native data set. After several rounds of refinement with a polyalanine model, sufficient side-chain density became apparent to assign a sequence register. Manual model building was done with COOT⁴⁹, and structure refinement was done using PHENIX⁵⁰ with four-fold NCS restraints. In later rounds of refinement, (GlcNAc)₂ molecules were modelled into the clear electron density features in the *F*_o - *F*_c map. In the final refined model, all four protomers contain the full-length ChbC except for the two N-terminus residues that are not resolved. The additional nine residues added during the cloning process are not resolved for three of the protomers, and partially resolved (435–438) in the fourth protomer. In addition, each asymmetric unit has four (GlcNAc)₂ molecules, four partially resolved NM molecules built with only the maltose head group, four NM molecules and six citrates.

Electrostatic potentials were calculated with DelPhi⁴¹ by solving the nonlinear Poisson-Boltzmann equation at physiological ionic strength (0.145 M). The calculations used a 1.4 Å probe radius, interior dielectric constant of 4, solvent dielectric constant of 80, and Debye-Hückel boundary conditions with a grid size of 251.

42. Punta, M. *et al.* Structural genomics target selection for the New York Consortium on Membrane Protein Structure. *J. Struct. Funct. Genomics* **10**, 255–268 (2009).
43. Levin, E. J., Quick, M. & Zhou, M. Crystal structure of a bacterial homologue of the kidney urea transporter. *Nature* **462**, 757–761 (2009).
44. Schaffner, W. & Weissmann, C. A rapid, sensitive, and specific method for the determination of protein in dilute solution. *Anal. Biochem.* **56**, 502–514 (1973).
45. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
46. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
47. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
48. Cowtan, K. Error estimation and bias correction in phase-improvement calculations. *Acta Crystallogr. D* **55**, 1555–1567 (1999).
49. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
50. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å

Yasufumi Umena^{1†*}, Keisuke Kawakami^{2†*}, Jian-Ren Shen² & Nobuo Kamiya^{1†}

Photosystem II is the site of photosynthetic water oxidation and contains 20 subunits with a total molecular mass of 350 kDa. The structure of photosystem II has been reported at resolutions from 3.8 to 2.9 Å. These resolutions have provided much information on the arrangement of protein subunits and cofactors but are insufficient to reveal the detailed structure of the catalytic centre of water splitting. Here we report the crystal structure of photosystem II at a resolution of 1.9 Å. From our electron density map, we located all of the metal atoms of the Mn₄CaO₅ cluster, together with all of their ligands. We found that five oxygen atoms served as oxo bridges linking the five metal atoms, and that four water molecules were bound to the Mn₄CaO₅ cluster; some of them may therefore serve as substrates for dioxygen formation. We identified more than 1,300 water molecules in each photosystem II monomer. Some of them formed extensive hydrogen-bonding networks that may serve as channels for protons, water or oxygen molecules. The determination of the high-resolution structure of photosystem II will allow us to analyse and understand its functions in great detail.

Photosystem II (PSII) is a membrane protein complex located in the thylakoid membranes of oxygenic photosynthetic organisms, and performs a series of light-induced electron transfer reactions leading to the splitting of water into protons and molecular oxygen. The products of PSII, namely chemical energy and oxygen, are indispensable for sustaining life on Earth. PSII from cyanobacteria is composed of 17 transmembrane subunits, three peripheral proteins and a number of cofactors, with a total molecular weight of 350 kDa. The light-induced oxidation of water is catalysed by a Mn₄Ca cluster that cycles through several different redox states (S_i, i = 0–4) on extraction of each electron by the PSII reaction centre, P₆₈₀^{1,2}. When four electrons and four protons are extracted from two molecules of water, one molecule of dioxygen is formed. The structure of PSII has been solved at resolutions from 3.8 to 2.9 Å in two closely related thermophilic cyanobacteria, *Thermosynechococcus elongatus*^{3–5} and *Thermosynechococcus vulcanus*⁶. These structural studies provided the arrangement of all of the protein subunits and the locations of chlorophylls and other cofactors, and formed a basis for further investigations into the functions of PSII. However, the resolution achieved so far is not high enough to reveal the structure of the Mn₄Ca cluster, the locations of substrate water molecules, or the precise arrangement of the amino-acid side chains and cofactors that may have significant mechanistic consequences for the energy, electron and proton transfer reactions. We have improved the resolution of the PSII crystals from *T. vulcanus* to 1.9 Å and analysed their structure (Methods, Supplementary Fig. 1 and Supplementary Table 1). This analysis provides many more details of the structure and of the coordination environments of the Mn₄Ca cluster and other cofactors than were previously available, and reveals the presence of a vast number of water molecules; these results may greatly advance our understanding of the energy, electron, proton transfer and water-splitting reactions taking place in PSII.

Overall structure

The overall structure is shown in Fig. 1. Every PSII monomer contains 19 protein subunits, among which PsbY was not found, suggesting

that this subunit has been lost during purification or crystallization, presumably owing to its loose association with PSII^{4–8}. The Cα superposition of our PSII dimer with the structure reported at a resolution of 2.9 Å (ref. 5) yielded a root mean square deviation of 0.78 Å, indicating that the overall structure determined at the lower resolution is well preserved in the present structure.

In addition to the protein subunits, there were 35 chlorophylls, two pheophytins, 11 β-carotenes, more than 20 lipids, two plastoquinones, two haem irons, one non-haem iron, four manganese atoms, three or four calcium atoms (one of which is in the Mn₄Ca cluster), three Cl[−] ions (two of which are in the vicinity of the Mn₄Ca cluster), one bicarbonate ion and more than 15 detergents in a monomer (Supplementary Table 2). Within each PSII monomer, more than 1,300 water molecules were found, yielding a total of 2,795 water molecules in the dimer (Fig. 1a and Supplementary Table 1). As shown in Fig. 1b, the water molecules were organized into two layers located on the surfaces of the stromal and luminal sides, respectively, with the latter having more water molecules than the former. A few water molecules were found within the membrane region, most of them serving as ligands to chlorophylls (see below). In the following, we describe the detailed structure and functions of the Mn₄CaO₅ cluster, as well as other cofactors, mainly on the basis of the structure of one of the two monomers, monomer A (chains labelled with capital letters in the accompanying Protein Data Bank file). There were some slight structural differences between the two monomers within the dimer; however, most of them are not related to the critical functions of PSII.

Structure of the Mn₄CaO₅ cluster

The electron densities of the four manganese atoms and the single calcium atom in the oxygen-evolving complex were well defined and clearly resolved, and the electron density for the calcium atom was lower than those of the manganese atoms, allowing us to identify the individual atoms unambiguously^{4,5} (Fig. 2a). In addition, five oxygen

¹Department of Chemistry, Graduate School of Science, Osaka City University, 3-3-138 Sugimoto, Sumiyoshi, Osaka 558-8585, Japan. ²Division of Bioscience, Graduate School of Natural Science and Technology/Faculty of Science, Okayama University, Okayama 700-8530, Japan. [†]Present addresses: Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan (Y.U.); Department of Chemistry, Graduate School of Science, Osaka City University, 3-3-138 Sugimoto, Sumiyoshi, Osaka 558-8585, Japan (K.K.); The OCU Advanced Research Institute for Natural Science and Technology (OCARINA), Osaka City University, 3-3-138 Sugimoto, Sumiyoshi, Osaka 558-8585, Japan (N.K.).

*These authors contributed equally to this work.

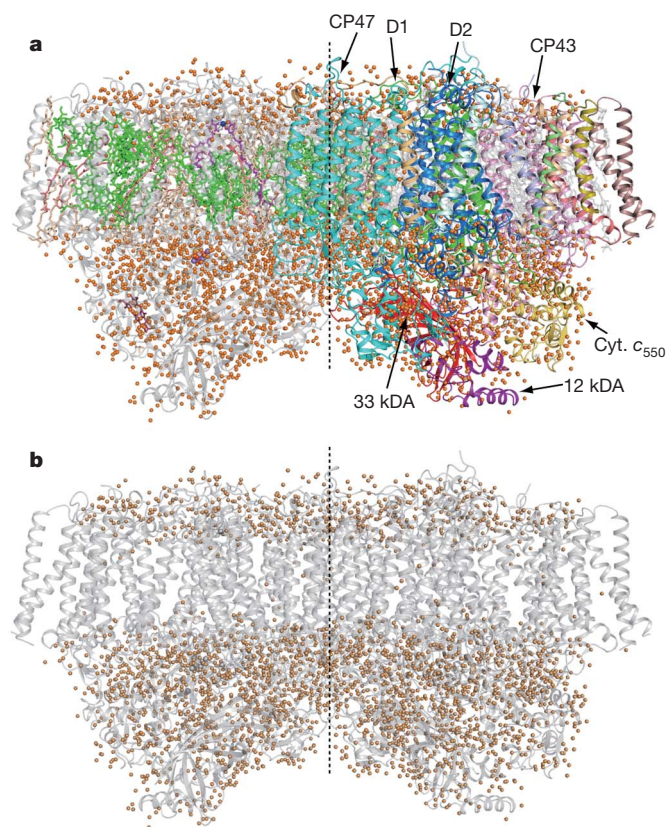


Figure 1 | Overall structure of PSII dimer from *T. vulcanus* at a resolution of 1.9 Å. View from the direction perpendicular to the membrane normal. **a**, Overall structure. The protein subunits are coloured individually in the right-hand monomer and in light grey in the left-hand monomer, and the cofactors are coloured in the left-hand monomer and in light grey in the right-hand monomer. Orange balls represent water molecules. **b**, Arrangement of water molecules in the PSII dimer. The protein subunits are coloured in light grey and all other cofactors are omitted. The central broken lines are the non-crystallographic two-fold axes relating the two monomers.

atoms were found to serve as oxo bridges linking the five metal atoms from the omit map (Fig. 2a). This gives rise to a Mn_4CaO_5 cluster. Of these five metals and five oxygen atoms, three manganese, one calcium and four oxygen atoms form a cubane-like structure in which the calcium and manganese atoms occupy four corners and the oxygen atoms occupy the other four. The bond lengths between the oxygens and the calcium in the cubane are generally in the range of 2.4–2.5 Å, and those between the oxygens and manganese are in the range of 1.8–2.1 Å (Fig. 2b). However, the bond length between one of the oxygens at the corner of the cubane (O5) and the calcium is 2.7 Å, and those between O5 and the manganese are in the range of 2.4–2.6 Å. Owing to these differences in bond lengths, the Mn_3CaO_4 cubane is not an ideal, symmetric one.

The fourth manganese (Mn4) is located outside the cubane and is linked to two manganese (Mn1 and Mn3) within the cubane by O5 and the fifth oxygen (O4) by a di- μ -oxo bridge. In this way, every two adjacent manganese are linked by di- μ -oxo bridges: Mn1 and Mn2 are linked by a di- μ -oxo bridge via O1 and O3, Mn2 and Mn3 are linked via O2 and O3, and Mn3 and Mn4 are linked via O4 and O5. The calcium is linked to all four manganese by oxo bridges: to Mn1 via the di- μ -oxo bridge formed by O1 and O5, to Mn2 via O1 and O2, to Mn3 via O2 and O5, and to Mn4 via the mono- μ -oxo bridge formed by O5. The whole structure of the Mn_4CaO_5 cluster resembles a distorted chair, with the asymmetric cubane serving as the seat base and the isolated Mn4 and O4 serving as the back of the chair. The

cubane-like structure has been reported previously^{4,9–12}, but the oxo bridges and exact distances among the individual atoms could not be determined at the medium resolution achieved previously⁴.

The distances among the four manganese determined for monomer A are 2.8 Å (Mn1–Mn2), 2.9 Å (Mn2–Mn3), 3.0 Å (Mn3–Mn4); 2.9 Å for monomer B, 3.3 Å (Mn1–Mn3) and 5.0 Å (Mn1–Mn4) (Fig. 2c). The distances between the calcium and the four manganese are 3.5 Å (Ca–Mn1), 3.3 Å (Ca–Mn2), 3.4 Å (Ca–Mn3) and 3.8 Å (Ca–Mn4) (Fig. 2d; for the corresponding distances in monomer B and the average distances between the two monomers, see Supplementary Table 3). These distances are largely different from those reported in the previous crystal structures^{3–6}; however, they are comparable to those reported from extended X-ray absorption fine structure studies^{13,14} if we consider that there is an error of 0.16 Å in the distances determined from the X-ray structural analysis (Methods).

In addition to the five oxygens, four water molecules (W1 to W4) were found to be associated with the Mn_4CaO_5 cluster, of which W1 and W2 are coordinated to Mn4 with respective distances of 2.1 and 2.2 Å, and W3 and W4 are coordinated to the calcium with a distance of 2.4 Å. No other water molecules were found to associate with the other three manganese, suggesting that some of the four waters may serve as the substrates for water oxidation.

All of the amino-acid residues coordinated to the Mn_4CaO_5 cluster were identified (Fig. 2e and Supplementary Table 4). Of these, D1-Glu 189 (D1 is one of the reaction centre subunits of PSII), served as a monodentate ligand to Mn1, which contradicts a previous report showing that it serves as a bidentate ligand⁵. All of the remaining five carboxylate residues served as bidentate ligands: D1-Asp 170 as a ligand to Mn4 and Ca, D1-Glu 333 to Mn3 and Mn4, D1-Asp 342 to Mn1 and Mn2, D1-Ala 344 (the carboxy-terminal residue of D1) to Mn2 and Ca, and CP43-Glu 354 to Mn2 and Mn3 (CP43 is one of the core antenna subunits of PSII). In addition, D1-His 332 is coordinated to Mn1, whereas D1-His 337 is not directly coordinated to the metal cluster. Most of the distances of the ligands to manganese are in the range of 2.0–2.3 Å; the two shortest distances are 1.9 Å, between D1-Glu 189 and Mn1, and 2.0 Å, between D1-Ala 344 and Mn2 (Supplementary Table 3). The distances of two carboxylate ligands to the calcium, D1-Asp 170 and D1-Ala 344, are slightly longer (2.3–2.4 Å) than the ligand distances to the manganese (Supplementary Table 3). Combining with the oxo bridges and waters, these give rise to a saturating ligand environment for the Mn_4CaO_5 cluster: each of the four manganese has six ligands whereas the calcium has seven ligands (Supplementary Table 4). The ligation pattern and the geometric positions of the metal atoms revealed in the present structure may have important consequences for the mechanisms of water splitting and O–O bond formation.

In addition to the direct ligands of the Mn_4CaO_5 cluster, we found that D1-Asp 61, D1-His 337 and CP43-Arg 357 are located in the second coordination sphere and may have important roles in maintaining the structure of the metal cluster, in agreement with various reports showing the importance of these three residues in maintaining the oxygen-evolving activity^{15–19}. One of the guanidinium η -nitrogens of CP43-Arg 357 is hydrogen-bonded to both O2 and O4 of the Mn_4CaO_5 cluster, whereas the other is hydrogen-bonded to the carboxylate oxygen of D1-Asp 170 and to that of D1-Ala 344. The imidazole ϵ -nitrogen of D1-His 337 is hydrogen-bonded to O3. These two residues may thus function to stabilize the cubane structure of the metal cluster as well as to provide partial positive charges to compensate for the negative charges induced by the oxo bridges and carboxylate ligands of the metal cluster. The carboxylate oxygen of D1-Asp 61 is hydrogen-bonded to W1, and also to O4 indirectly through another water molecule, suggesting that this residue may also contribute to stabilizing the metal cluster. Furthermore, D1-Asp 61 is located at the entrance of a proposed proton exit channel involving a chloride ion (Cl^- ; see below), suggesting that this residue may function in facilitating proton exit from the Mn_4CaO_5 cluster^{5,20,21}.

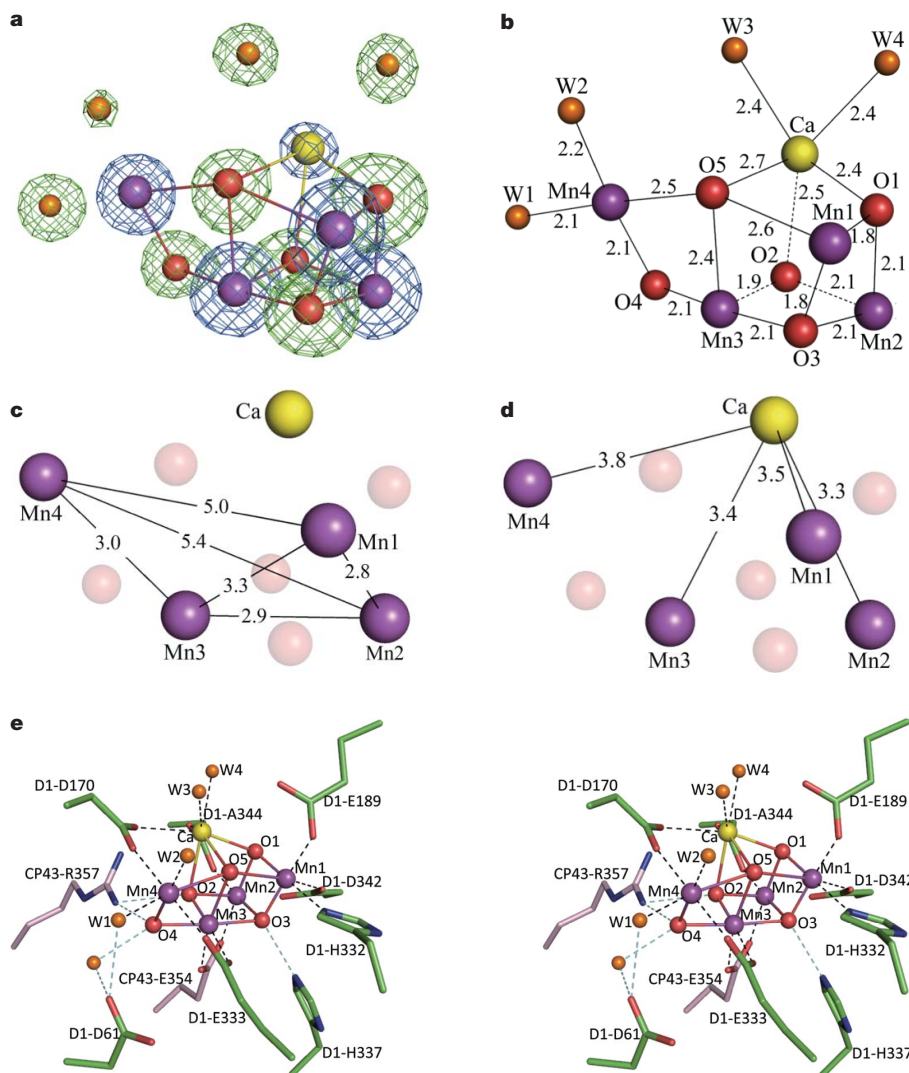


Figure 2 | Structure of the Mn_4CaO_5 cluster. **a**, Determination of individual atoms associated with the Mn_4CaO_5 cluster. The structure of the cluster was superimposed with the $2F_o - F_c$ map (blue) contoured at 5σ for manganese and calcium atoms, and with the omit map (green) contoured at 7σ for oxygen atoms and water molecules. **b**, Distances (in Å) between metal atoms

and oxo bridges or water molecules. **c**, Distances between each pair of manganese atoms. **d**, Distances between the manganese and the calcium atoms. **e**, Stereo view of the Mn_4CaO_5 cluster and its ligand environment. The distances shown are the average distances between the two monomers. Manganese, purple; calcium, yellow; oxygen, red; D1, green; CP43, pink.

The most significant structural feature of the Mn_4CaO_5 cluster, which may be important for elucidating the mechanism of the water-splitting reaction, is its distorted chair form. The large distortion is principally caused by the existence of the calcium and O5 in the Mn_4CaO_5 cluster, as described above. The apparently longer distances between O5 and metal atoms suggested that the corresponding bonds are weak, and that O5 may therefore have a lower negative charge than the valence of -2 expected for normal oxygen atoms in oxo bridges. This in turn suggests that O5 may exist as a hydroxide ion in the S_1 state and may provide one of the substrates for dioxygen formation. Because both W2 and W3 are within the hydrogen-bond distances to O5, one of these two waters may provide another substrate.

Because the transition between S_0 and S_1 is fastest in the Kok cycle, the proton released during this transition may be accepted by D1-Tyr 161 (also termed Y_Z), which is deprotonated by means of proton-coupled electron transfer (PCET; see below). W3 is closer to Y_Z than is O5 (Fig. 3a) and may be a more favourable candidate than O5 as the proton-releasing group. Thus, W3 rather than O5 may be a hydroxide ion in the S_1 state, suggesting that O–O bond formation may occur between W2 and W3. In any case, our results suggest that the O–O bond formation occurs in two of the three species O5, W2 and W3.

Hydrogen-bond network around Y_Z

Y_Z is located between the Mn_4CaO_5 cluster and the PSII reaction centre, and functions to mediate electron transfer between the two. We found an extensive hydrogen-bonding network between Y_Z and the Mn_4CaO_5 cluster and from Y_Z to the lumenal bulk phase. Y_Z was hydrogen-bonded to the two waters coordinated to the calcium either directly (W4) or indirectly through another water (W3; Fig. 3a). The hydrogen bond between the additional water and Y_Z that mediates the link from W3 to Tyr161 has a length of 2.6 \AA , suggesting that this is a strong (low-barrier) hydrogen bond²². This additional water also mediates the hydrogen bond between the two waters bound to Mn4 and Y_Z . Furthermore, another strong hydrogen bond was found between Y_Z and the ϵ -nitrogen of D1-His 190, which is 2.5 \AA in length and lies on the opposite side of the Mn_4CaO_5 cluster. D1-His 190 was further hydrogen-bonded to D1-Asn 298 and to several waters and residues including CP43-Ala 411, D1-Asn 322 and PsbV-Tyr 137 (the C-terminal residue of the PsbV subunit), leading to an exit pathway to the lumenal bulk solution (Fig. 3b). This hydrogen-bond network is located in the interfaces between the D1, CP43 and PsbV subunits and may function as an exit channel for protons that arise from PCET via Y_Z . This provides support for the existence of a PCET pathway involving Y_Z .

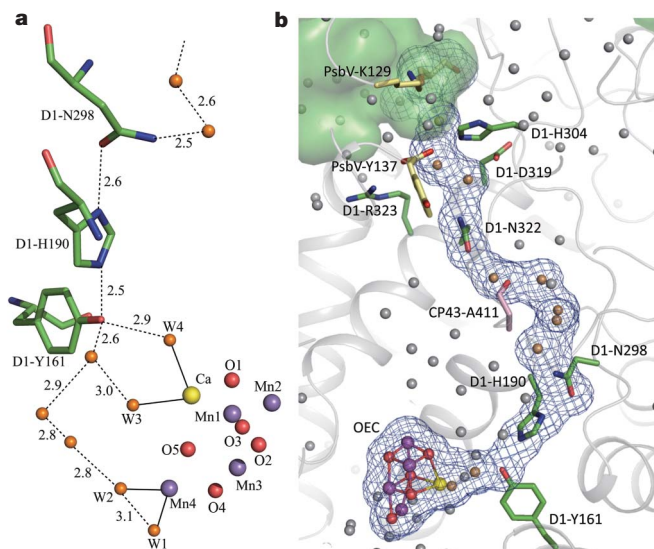


Figure 3 | Hydrogen-bond network around Y_Z. **a**, Hydrogen bonds around Y_Z (D1-Tyr 161). The bonds between metal atoms and water ligands are depicted as solid lines, and the hydrogen bonds are depicted as dashed lines. Distances are expressed in ångströms. **b**, Hydrogen-bond network from the Mn₄CaO₅ cluster through Y_Z to the lumenal bulk phase. Water molecules participating in the hydrogen-bond network are depicted in orange, whereas those not participating are depicted in grey. The area in green in the upper left corner represents the lumenal bulk surface. PsbV, pale yellow; other colour codes are the same as in Fig. 2. OEC, oxygen-evolving complex.

and D1-His 190, as implied by a number of previous studies^{23–25}. PsbV-Tyr 137, at the exit of this channel, is surrounded by several charged residues including D1-Arg 323, D1-His 304 and PsbV-Lys 129; these residues may therefore function to regulate the proton excretion through the PCET pathway (Fig. 3b).

The other redox-active tyrosine residue, Y_D (D2-Tyr 160), has a different, rather hydrophobic, environment from that of Y_Z. For a discussion of the environment of Y_D, see Supplementary Fig. 3 and discussions.

The structure and function of chloride-binding sites

Previous studies have identified two chloride ions (Cl[−]) in the vicinity of the Mn₄CaO₅ cluster by substitution of Br[−] or I[−] for Cl[−] (refs 26, 27), although only one Cl[−] site was visible in the native PSII crystals⁵. In the present study, the electron density for the two Cl[−]-binding sites were clearly visible (Fig. 4a), which were confirmed from the anomalous difference Fourier map calculated with data collected at a wavelength of 1.75 Å (Fig. 4a). The two Cl[−]-binding sites are located in the same position as those reported for Br[−]- or I[−]-substituted PSII previously^{26,27} (Fig. 4b, c). Both Cl[−] ions are surrounded by four species, among which two are waters. For one of the ions, Cl[−]1, the other two species are the amino group of D2-Lys 317 and the backbone nitrogen of D1-Glu 333, and for the other ion, Cl[−]2, they are the backbone nitrogens of D1-Asn 338 and CP43-Glu 354. Because the side chains of D1-Glu 333 and CP43-Glu 354 are coordinated to the Mn₄CaO₅ cluster directly, the two Cl[−] anions may function to maintain the coordination environment of the Mn₄CaO₅ cluster, thereby allowing the oxygen-evolving reaction to proceed properly.

In addition to the structural roles, the two Cl[−]-binding sites were found to lie at the entrance of hydrogen-bond networks starting from the Mn₄CaO₅ cluster and extending towards the lumenal bulk solution (Fig. 4b, c). The network through Cl[−]1 was located in the interface of the D1, D2 and PsbO subunits, and that through Cl[−]2 was located in the interface of the D1, CP43 and PsbU subunits. These hydrogen-bond networks involve a number of bound waters and

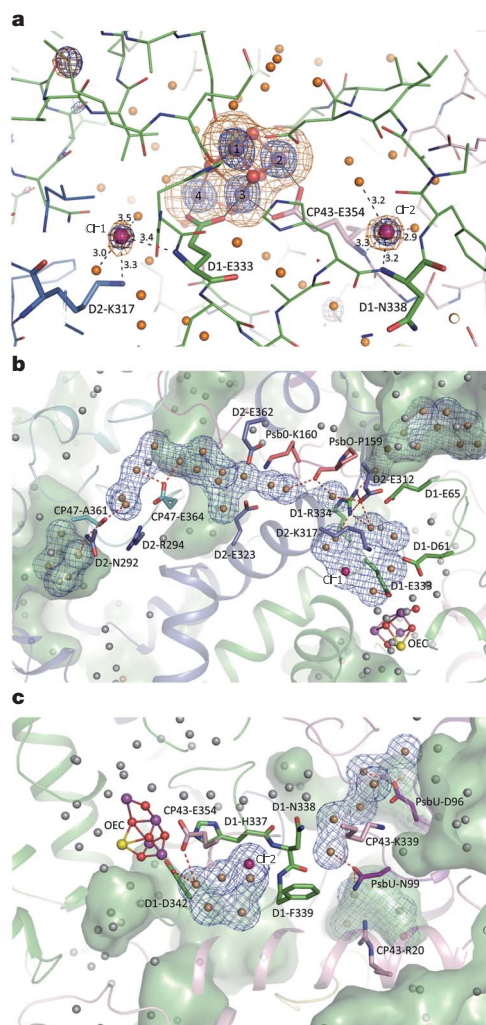


Figure 4 | Structure of two Cl[−]-binding sites. **a**, Location of the two Cl[−] ions. Blue mesh, 2F_o − F_c map contoured at 4σ, measured at a wavelength of 0.9 Å; orange mesh, anomalous difference Fourier map contoured at 8σ, measured at a wavelength of 1.75 Å. The small density at the upper left corner is from the sulphur atom of D1-Met 328. Distances are expressed in ångströms. **b**, Hydrogen-bond network from the Mn₄CaO₅ cluster through the Cl[−]1-binding site to the lumenal bulk phase. **c**, Hydrogen-bond network from the Mn₄CaO₅ cluster through the Cl[−]2-binding site to the lumenal bulk phase. Colour codes are the same as in Figs 2 and 3.

some hydrophilic or charged amino-acid residues; they thus may function as either proton exit channels or water inlet channels.

Chlorophylls and β-carotenes

The positions and orientations of most chlorophylls are similar to those reported previously^{4–6}. However, we determined the ligands to the central magnesium of all of chlorophylls, of which seven are coordinated by water instead of amino-acid residues (Fig. 5a and Supplementary Table 5). These are Chl6 (the accessory chlorophyll of D1); Chl7 (the accessory chlorophyll of D2); Chl12, Chl18 and Chl21, harboured by CP47; and Chl-31 and Chl-34, harboured by CP43. In addition, Chl38 was coordinated by CP43-Asn 39 and all other chlorophylls are coordinated by histidines. From our electron density map, we confirmed that all of the C8 and C13 positions in the phytol chains have a (*R, R*) configuration, in agreement with the stereochemistry determined for the complete phytol chain^{28,29}. Furthermore, we found that most of the vinyl groups are located in or near the same plane of the tetrapyrrole ring, which may contribute to the extension of energy coupling within the plane and hence facilitate the energy migration between adjacent chlorophylls.

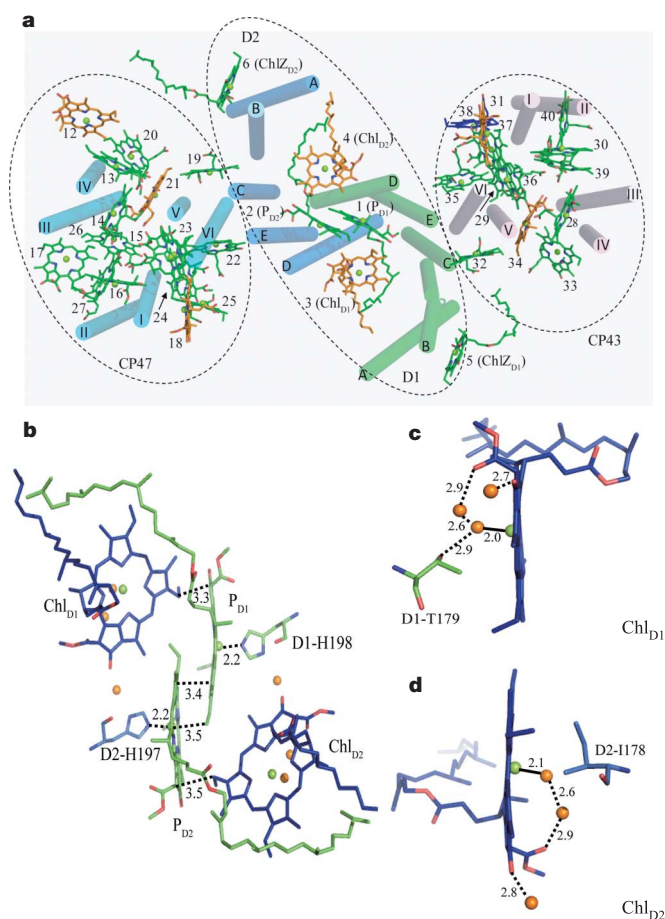


Figure 5 | Organization of chlorophylls. **a**, Organization of 35 chlorophylls in a PSII monomer. Chlorophylls (Chl) whose central magnesium atoms are coordinated by water are depicted in orange, and one chlorophyll coordinated by CP43-Asn 39 is depicted in blue. All other chlorophylls are coordinated by His and are depicted in green. Transmembrane helices of D1 and D2 are labelled A–E, and transmembrane helices of CP47 and CP43 are labelled I–VI. **b**, Organization of the four reaction-centre chlorophylls. Magnesium atoms of chlorophylls are depicted in green, and water molecules are depicted in orange. The edge-to-edge distances are expressed in Ångströms. **c**, Water ligand and hydrogen bonds of Chl_{D1}. **d**, Water ligand and hydrogen bonds of Chl_{D2}.

The four chlorophylls constituting the PSII reaction centre are depicted in Fig. 5b. The non-crystallographic two-fold symmetry expected for the chlorophyll dimer comprising P_{D1} and P_{D2} seems to be broken in our high-resolution structure, as follows. The vinyl group of P_{D1} is roughly in plane, and its terminal carbon atom is close to the magnesium of P_{D2} at its sixth coordination site. In contrast, the corresponding vinyl group of P_{D2} is out of the chlorin plane and is located some distance from the magnesium of P_{D1}. The edge-to-edge distances that are able to form π – π stacking or CH– π stacking among the four chlorophylls range from 3.3 to 3.5 Å, with the shortest being 3.3 Å, between P_{D1} and Chl_{D1}. This may account partly for the preferential electron transfer along the D1 side. Importantly, although both water ligands to the two ‘accessory chlorophylls’ Chl_{D1} and Chl_{D2} are hydrogen-bonded to the carbonyl oxygen of the methoxycarbonyl group of chlorin ring V, the water ligand of Chl_{D1} is further hydrogen-bonded to D1-Thr 179 but no such a hydrogen-bond partner is found for Chl_{D2} (Fig. 5c, d). These discrepancies may also contribute to the functional differences between the two chlorophylls.

CP47 and CP43 bind 16 and 13 antenna chlorophylls, respectively, which are arranged as double layers connected by a special chlorophyll at the middle of the two layers³⁰. The chlorophylls are distributed in one of three areas separated by coiled-coil helix dimers (I, II) and

(V, VI) of CP47 (or CP43) (Fig. 5a, Supplementary Fig. 4 and Supplementary Discussions). A significant feature of the present structure is that the chlorin rings of most of the chlorophylls are not planar, which may affect the electronic, spectroscopic or energetic properties of the chlorophylls.

The positions and orientations of most of the β -carotenes are similar to those in the previous structures (Supplementary Fig. 5 and Supplementary Discussions). Each of the 11 β -carotenes was found to be of all-*trans* type.

Plastoquinones, non-haem iron and lipids

The two plastoquinones, Q_A and Q_B, were found in positions similar to those reported previously (Supplementary Fig. 5), with Q_B being less defined and having a higher B-factor than Q_A. Q_C, a third plastoquinone found in the previous structure⁵, was not found in the present structure, probably as a result of the differences in preparation or crystallization conditions between the previous and present studies.

Six monogalactosyldiacylglycerol (MGDG), five digalactosyldiacylglycerol (DGDG), four sulfoquinovosyldiacylglycerol (SQDG) and five phosphatidylglycerol molecules were found (Supplementary Table 2, Supplementary Fig. 5 and Supplementary Discussions). All of the SQDGs and phosphatidylglycerols were distributed in the stromal side, with their head groups located in the stromal surface of the membrane, whereas all of the MGDGs and DGDGs, except for one MGDG, were located in the luminal side. This may suggest that the hydrophilic head groups of SQDG and phosphatidylglycerol cannot penetrate the membrane, resulting in their preferential distribution in the stromal side, but that the more hydrophobic lipids MGDG and DGDG were able to transfer across the membrane.

For the structure of the non-haem iron and bicarbonate, see Supplementary Fig. 6 and Supplementary Discussions.

Conclusion

The high-resolution structure of PSII reveals the geometric arrangement of the Mn₄CaO₅ cluster as well as its oxo bridges and ligands, and four bound water molecules. This provides a basis for unravelling the mechanism of water splitting and O–O bond formation, one of nature’s most fascinating and important reactions. In addition, our determination of the precise arrangement of amino-acid side chains and cofactors gives us a solid structural understanding of energy migration, electron transfer and water-splitting reactions taking place within PSII.

METHODS SUMMARY

We purified PSII core complexes highly active in oxygen evolution from a thermophilic cyanobacterium, *T. vulcanus*^{31,32}. The homogeneity of PSII was improved by introducing a re-crystallization step. Previous crystallization conditions^{6,32} were improved to produce high-resolution crystals (Methods). A typical diffraction pattern is shown in Supplementary Fig. 1, from which diffraction spots beyond a resolution of 1.8 Å could be observed. To suppress the possible radiation damage to a minimum level, we used a slide-oscillation method, resulting in the X-ray dose at each point of the crystal being lower than in previous experiments³³. A full data set was collected at a wavelength of 0.9 Å and processed to a resolution of 1.9 Å (Supplementary Table 1). For identifying the positions of Cl[−] ions, another data set was taken at a wavelength of 1.75 Å, and processed to a resolution of 2.5 Å.

The structure of PSII was solved by the molecular replacement method using the structure reported at a resolution of 2.9 Å as the search model⁵ (Protein Data Bank ID, 3BZ1), and refined to R_{cryst} and R_{free} values of 0.174 and 0.201, respectively, with a Cruickshank diffraction-component precision index³⁴ of 0.11 Å. Detailed procedures for crystallization and structure determination can be found in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 December 2010; accepted 8 February 2011.

Published online 17 April 2011.

1. Kok, B., Forbush, B. & McGloin, M. Cooperation of charges in photosynthetic oxygen evolution. I. A linear four step mechanism. *Photochem. Photobiol.* **11**, 457–475 (1970).

2. Joliot, P. Period-four oscillations of the flash-induced oxygen formation in photosynthesis. *Photosynth. Res.* **76**, 65–72 (2003).
3. Zouni, A. *et al.* Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution. *Nature* **409**, 739–743 (2001).
4. Ferreira, K. N., Iverson, T. M., Maghlaoui, K., Barber, J. & Iwata, S. Architecture of the photosynthetic oxygen-evolving center. *Science* **303**, 1831–1838 (2004).
5. Guskov, A. *et al.* Cyanobacterial photosystem II at 2.9 Å resolution and role of quinones, lipids, channels and chloride. *Nature Struct. Mol. Biol.* **16**, 334–342 (2009).
6. Kamiya, N. & Shen, J.-R. Crystal structure of oxygen-evolving photosystem II from *Thermosynechococcus vulcanus* at 3.7-Å resolution. *Proc. Natl Acad. Sci. USA* **100**, 98–103 (2003).
7. Kawakami, K., Iwai, M., Ikeuchi, M., Kamiya, N. & Shen, J.-R. Location of PsbY in oxygen-evolving photosystem II revealed by mutagenesis and X-ray crystallography. *FEBS Lett.* **581**, 4983–4987 (2007).
8. Broser, M. *et al.* Crystal structure of monomeric photosystem II from *Thermosynechococcus elongatus* at 3.6 Å resolution. *J. Biol. Chem.* **285**, 26255–26262 (2010).
9. De Paula, J. C., Beck, W. F. & Brudvig, G. W. Magnetic properties of manganese in the photosynthetic O₂-evolving complex. 2. Evidence for a manganese tetramer. *J. Am. Chem. Soc.* **108**, 4002–4009 (1986).
10. Carrell, G., Tyrishkin, A. M., & Dismukes, G. C. An evaluation of structural models for the photosynthetic water-oxidizing complex derived from spectroscopic and X-ray diffraction signatures. *J. Biol. Inorg. Chem.* **7**, 2–22 (2002).
11. Vincent, J. B. & Christou, G. A molecular ‘double-pivot’ mechanism for water oxidation. *Inorg. Chim. Acta* **136**, L41–L43 (1987).
12. Peloquin, J. M. & Britt, R. D. EPR/ENDOR characterization of the physical and electronic structure of the OEC Mn cluster. *Biochim. Biophys. Acta* **1503**, 96–111 (2001).
13. Robblee, J. H., Cince, R. M. & Yachandra, V. K. X-ray spectroscopy-based structure of the Mn cluster and mechanism of photosynthetic oxygen evolution. *Biochim. Biophys. Acta* **1503**, 7–23 (2001).
14. Zein, S. *et al.* Focusing the view on nature’s water-splitting catalyst. *Phil. Trans. R. Soc. B* **363**, 1167–1177 (2008).
15. Nixon, P. J. & Diner, B. Analysis of water-oxidation mutants constructed in the cyanobacterium *Synechocystis* sp. PCC 6803. *Biochem. Soc. Trans.* **22**, 338–343 (1994).
16. Chu, H.-A., Nguyen, A. P. & Debus, R. J. Amino acid residues that influence the binding of manganese or calcium to Photosystem II. 1. The luminal inter-helical domains of the D1 polypeptide. *Biochemistry* **34**, 5839–5858 (1995).
17. Hwang, H. J., Dilbeck, P., Debus, R. J. & Burnap, R. L. Mutation of arginine 357 of the CP43 protein of photosystem II severely impairs the catalytic S-state cycle of the H₂O oxidation complex. *Biochemistry* **46**, 11987–11997 (2007).
18. Debus, R. J. Protein ligation of the photosynthetic oxygen-evolving center. *Coord. Chem. Rev.* **252**, 244–258 (2008).
19. Service, R. J., Hillier, W. & Debus, R. J. Evidence from FTIR difference spectroscopy of an extensive network of hydrogen bonds near the oxygen-evolving Mn₄Ca cluster of photosystem II involving D1-Glu65, D2-Glu312, and D1-Glu329. *Biochemistry* **49**, 6655–6669 (2010).
20. Murray, J. W. & Barber, J. Structural characteristics of channels and pathways in photosystem II including the identification of an oxygen channel. *J. Struct. Biol.* **159**, 228–237 (2007).
21. Ho, F. M. & Styring, S. Access channels and methanol binding site to the CaMn₄ cluster in Photosystem II based on solvent accessibility simulation, with implications for substrate water access. *Biochim. Biophys. Acta* **1777**, 140–153 (2008).
22. Zhang, C. Low-barrier hydrogen bond plays key role in active photosystem II—A new model for photosynthetic water oxidation. *Biochim. Biophys. Acta* **1767**, 493–499 (2007).
23. Hoganson, C. W. & Babcock, G. T. A metalloradical mechanism for the generation of oxygen from water in photosynthesis. *Science* **277**, 1953–1956 (1997).
24. Tommos, C. & Babcock, G. T. Proton and hydrogen currents in photosynthetic water oxidation. *Biochim. Biophys. Acta* **1458**, 199–219 (2000).
25. Hays, A.-M. A., Vassiliev, I. R., Golbeck, J. H. & Debus, R. J. Role of D1-His190 in the proton-coupled oxidation of tyrosine Y₂ in manganese-depleted photosystem II. *Biochemistry* **38**, 11851–11865 (1999).
26. Murray, J. W. *et al.* X-ray crystallography identifies two chloride binding sites in the oxygen evolving centre of photosystem II. *Energy Environ. Sci.* **1**, 161–166 (2008).
27. Kawakami, K., Umena, Y., Kamiya, N. & Shen, J.-R. Location of chloride and its possible functions in oxygen-evolving photosystem II revealed by X-ray crystallography. *Proc. Natl Acad. Sci. USA* **106**, 8567–8572 (2009).
28. Burrell, J. W. K., Jackman, L. M. & Weedon, B. L. C. Stereochemistry and synthesis of phytol, geraniol, and nerol. *Proc. Chem. Soc.* **1959**, 263–264 (1959).
29. Crabbe, P., Djerassi, C., Eisenbraun, E. J. & Liu, S. Optical rotatory dispersion studies. XXIX. Absolute configuration of phytol. *Proc. Chem. Soc.* **1959**, 264–265 (1959).
30. Vasil’ev, S., Orth, P., Zouni, A., Owens, T. G. & Bruce, D. Excited-state dynamics in photosystem II: Insights from the x-ray crystal structure. *Proc. Natl Acad. Sci. USA* **98**, 8602–8607 (2001).
31. Shen, J.-R. & Inoue, Y. Binding and functional properties of two new extrinsic components, cytochrome c-550 and a 12 kDa protein, in cyanobacterial photosystem II. *Biochemistry* **32**, 1825–1832 (1993).
32. Shen, J.-R. & Kamiya, N. Crystallization and the crystal properties of the oxygen-evolving photosystem II from *Synechococcus vulcanus*. *Biochemistry* **39**, 14739–14744 (2000).
33. Yano, J. *et al.* X-ray damage to the Mn₄Ca complex in single crystals of photosystem II: a case study for metalloprotein crystallography. *Proc. Natl Acad. Sci. USA* **102**, 12047–12052 (2005).
34. Cruickshank, D. W. J. Remarks about protein structure precision. *Acta Crystallogr. D* **55**, 583–601 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The X-ray diffraction data was taken at beamlines BL44XU, BL41XU and BL38B1 at SPring-8. We thank E. Yamashita, N. Shimizu, S. Baba and N. Mizuno for their help in using the beamlines. J.-R.S. thanks Y. Inoue for his support in the initiation of this work. This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas (Structures of Biological Macromolecular Assemblies), a Grant-in-Aid for Creative Scientific Research, a GCOE programme on Pico-biology at the University of Hyogo, a Grant-in-Aid for Scientific Research (C), from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and a research grant from the Yamada Science foundation.

Author Contributions K.K. performed, and J.-R.S. supervised, the purification and crystallization of PSII. K.K., Y.U. and J.-R.S. performed X-ray diffraction experiments. Y.U. analysed the structure, and N.K. supervised the structure analysis and refinement process. J.-R.S. and N.K. jointly wrote the paper, and all of the authors joined the discussion of the results.

Author Information Atomic coordinates have been deposited in the Protein Data Bank under the accession number 3ARC. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.-R.S. (shen@cc.okayama-u.ac.jp) and N.K. (nkamiya@sci.osaka-cu.ac.jp).

METHODS

Purification and crystallization. Highly active, dimeric PSII was purified from the thermophilic cyanobacterium *T. vulcanus* following refs 31, 32, and the crystals were grown as described previously^{6,32}. To improve the crystal quality, the purity and homogeneity of PSII was improved by introducing a re-crystallization step in which the PSII core complexes were first crystallized in 12–24 h on ice or at 4 °C, and the microcrystals obtained were collected, re-solubilized and used for the second crystallization step. Sometimes the re-crystallization step was repeated to ensure a higher homogeneity of the samples, which were monitored by dynamic light-scattering measurements. The re-crystallization procedure typically decreased the polydispersity of the samples from 30% to around 20%.

The PSII crystals obtained were subjected to a post-crystallization dehydration procedure by increasing the concentrations of glycerol and PEG in the following way. The crystals were first transferred into a 150- μ l buffer solution containing 6% PEG 3000 in place of 4–5% PEG 1450 in the original crystallization buffer concentrations (which contained no glycerol). After 25–30 min of incubation at 12 °C, half of the buffer volume was replaced with a new buffer containing a 1.4%-higher concentration of PEG 3000 and an additional 2.5% glycerol. This procedure was repeated every 25–30 min until the concentrations of glycerol and PEG 3000 reached 25% and 20%, respectively, in the final buffer. The crystals were then dehydrated by evaporation against air with a humidity of 75–90% in the final buffer for 2.5 h in the incubator at 12 °C, frozen in flash-cooling nitrogen gas and stored in liquid nitrogen. The crystals thus obtained had an approximate water content of 57%, which is much lower than that of the crystals obtained previously³² (66%) and that of the crystals used to analyse the structure at a resolution of 2.9 Å (ref. 5; 61%). All of the cryoprotectant replacement and cryocooling procedures were carried out under dim green light to avoid possible advancement of the S-states in the Kok cycle. A typical diffraction pattern of the PSII crystals is shown in Supplementary Fig. 1, in which diffraction spots beyond a resolution of 1.8 Å can be observed.

Data collection. After dehydration, the crystals were coated with a mixture of oil containing 66.5% Paratone-N, 28.5% paraffin oil and 5% glycerol, and flash-cooled at 100 K with a nitrogen gas stream. Two diffraction data sets were collected from two PSII crystals, one with a wavelength of 0.9 Å and the other one with a wavelength of 1.75 Å, at beamline BL44XU of SPring-8 (Japan). The X-ray beam had a size of 50 \times 50 μ m², and the diffraction images were recorded with a Mar225HE charge-coupled-device detector. For the data set taken at 0.9 Å, we used a large PSII crystal with a size of 0.2 \times 0.7 \times 1.0 mm³. The crystal was shifted by 30 μ m to an adjacent point along the oscillation axis after recording 100 oscillation images, each of which was rotated by 0.2° relative to the last. Each point therefore covered a range of 20°. We collected a total of 900 images from nine irradiation points, covering a rotation angle of 180°. The data were processed and scaled using XDS and XSCALE³⁵ (Supplementary Table 1).

The photon flux of the beamline used (BL44XU) was 0.7×10^{11} photons s⁻¹ (with an attenuator of 0.2-mm aluminium), and the exposure time was 1 s for each diffraction image. This gave rise to a total X-ray dose of 2.5×10^{10} photons μ m⁻² for the total of 900 images. Because the whole data set was divided into nine spots on the crystal, each spot received a total dose of 0.28×10^{10} photons μ m⁻². If we consider that each point was rotated by 20° during data collection, the X-ray dose on a unit volume of the crystal will be slightly lower. This dose is much lower than that used previously^{3–6}, and is also at a low level of the dose range reported to induce possible radiation damage in the Mn₄CaO₅ cluster³³.

For the data set taken at a wavelength of 1.75 Å, we collected 2,400 oscillation images, each rotated by 0.3° over a range of 360°. For each of the data wedges of 10°, an inverse beam geometry was used to measure the Friedel pairs directly. The data was processed with HKL2000³⁶, and the reflection data statistics are summarized in Supplementary Table 1.

Structure refinement. An initial structure of the PSII dimer was obtained with the molecular replacement method of CNS³⁷ using the structure of PSII monomer⁵ (PDB ID, 3BZ1) as a search model. The first stage of structure refinement was carried out using the CNS program package and the second stage was performed with REFMAC5 in the CCP4 program suite³⁸. The two monomers in the PSII dimer were refined separately, and the structural model was revised using COOT³⁹. Structures of cofactors, lipids, detergents and water molecules were determined and refined as described below. The refinement statistics are presented in Supplementary Table 1.

Mn₄CaO₅ cluster. The locations of the metal atoms of the Mn₄CaO₅ cluster, namely four manganese atoms and one calcium atom, were determined using a

composite omit $2F_o - F_c$ map. Oxygen atoms forming oxo bridges in the Mn₄Ca cluster were identified and determined with an $F_o - F_c$ omit map. The electron density of the O5 atom in the Mn₄CaO₅ cluster was affected heavily by electron density distributions of the nearby metal atoms, which interfered with the determination of its location in the $F_o - F_c$ omit map. Thus, the position of O5 was determined from the $2F_o - F_c$ map. The average *B*-factor of the five metal ions refined without restraint was 25.3 Å², which was lower than that of the overall average *B*-factor, of 35.2 Å² (Supplementary Table 1).

Chloride and metal ions. The existence of two Cl⁻ ions in the vicinity of the oxygen-evolving complex was previously reported with Br⁻ or I⁻-substituted PSII^{26,27}. We confirmed the positions of the two Cl⁻-binding sites in native PSII both with an $F_o - F_c$ omit map taken at a wavelength of 0.9 Å and analysed to a resolution of 1.9 Å, and with an anomalous difference Fourier map taken at a wavelength of 1.75 Å and analysed to a resolution of 2.5 Å. Several additional calcium and magnesium ions were identified by these two electron density maps, and their structures were constructed by taking their coordination environments into consideration.

Chlorophyll *a* and pheophytin molecules. Electron density distributions for the magnesium atoms of chlorophyll *a* were clearly separated from those for the chlorin rings and were located out of the ring planes in most cases. The chlorin rings were bent to various degrees depending on their environments. The conformations of ethyl and vinyl groups were determined unambiguously from the corresponding electron density distributions. Two optically active centres (C8 and C13) of all of the phytol chains were also recognized as being in the (*R*, *R*) configuration from the electron density map.

Plastoquinones. Two plastoquinones, Q_A and Q_B, were identified from the electron density map, whereas a third plastoquinone, Q_C, reported in a previous structure⁵, was not observed. Q_A had a well-defined electron density distribution, resulting in a low average *B*-factor of 25.5 Å², whereas the electron density for Q_B was weak, resulting in a higher *B*-factor of 76.8 Å².

Lipids and unknown molecules. Two kinds of lipid molecule, SQDG and phosphatidylglycerol, contained sulphur and phosphorous atoms, respectively, which have larger anomalous dispersion effects at a longer wavelength. The positions of four of eight SQDG molecules and eight of ten phosphatidylglycerol molecules in the PSII dimer were confirmed from the anomalous dispersion of the sulphur and phosphate atoms contained in these lipids, on the basis of the anomalous difference Fourier map calculated from the data set taken at a wavelength of 1.75 Å. The electron densities for a typical SQDG and phosphatidylglycerol are depicted in Supplementary Fig. 2. Other lipid molecules were found and modelled on the basis of the $F_o - F_c$ omit map and the $2F_o - F_c$ map. Six lipids with two fatty-acid chains were found in the dimer, but their species could not be identified. Additionally, 30 single alkyl chains of unknown identity were observed in the dimer; of these, 23 were located adjacently. Therefore, the total number of lipids should exceed 23 in each monomer.

Water molecules. Water molecules were assigned from the $2F_o - F_c$ electron density map at over the 1 σ level. Around 1,300 water molecules were found in each monomer (Supplementary Table 1), and a few of them were found to be disordered.

Error estimation for atomic coordinates. The coordinate error was estimated with the diffraction-component precision index (DPI) introduced by Cruickshank^{34,40}, using the software SFCHECK in the CCP4 suite³⁸. The DPI value of the whole PSII structure was found to be 0.11 Å, resulting in a standard uncertainty in the bond length of 0.16 Å.

- Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
- Otwinowski, Z. & Minor, M. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Daopin, S., Davies, D. R., Schlunegger, M. P. & Grütter, M. G. Comparison of two crystal structures of TGF- β 2: the accuracy of refined protein structures. *Acta Crystallogr. D* **50**, 85–92 (1994).

Single-ion quantum lock-in amplifier

Shlomi Kotler¹, Nitzan Akerman¹, Yinnon Glickman¹, Anna Keselman¹ & Roei Ozeri¹

Quantum metrology¹ uses tools from quantum information science to improve measurement signal-to-noise ratios. The challenge is to increase sensitivity while reducing susceptibility to noise, tasks that are often in conflict. Lock-in measurement is a detection scheme designed to overcome this difficulty by spectrally separating signal from noise. Here we report on the implementation of a quantum analogue to the classical lock-in amplifier. All the lock-in operations—modulation, detection and mixing—are performed through the application of non-commuting quantum operators to the electronic spin state of a single, trapped Sr^+ ion. We significantly increase its sensitivity to external fields while extending phase coherence by three orders of magnitude, to more than one second. Using this technique, we measure frequency shifts with a sensitivity of $0.42 \text{ Hz Hz}^{-1/2}$ (corresponding to a magnetic field measurement sensitivity of $15 \text{ pT Hz}^{-1/2}$), obtaining an uncertainty of less than 10 mHz (350 fT) after 3,720 seconds of averaging. These sensitivities are limited by quantum projection noise and improve on other single-spin probe technologies^{2,3} by two orders of magnitude. Our reported sensitivity is sufficient for the measurement of parity non-conservation⁴, as well as the detection of the magnetic field of a single electronic spin one micrometre from an ion detector with nanometre resolution. As a first application, we perform light shift spectroscopy of a narrow optical quadrupole transition. Finally, we emphasize that the quantum lock-in technique is generic and can potentially enhance the sensitivity of any quantum sensor.

Quantum probes with unprecedented sensitivities are advancing the field of metrology. In particular, cold, trapped ions are well isolated from their environment and their internal states and motion can be controlled with high fidelity, thus enabling researchers to use them as excellent probes^{5,6}.

Achieving a high signal-to-noise ratio involves demands—decreasing the effect of noise on the probe while enhancing its response to the measured signal—that are often in conflict. The problem arises if the noise and the signal couple to the probe through the same physical channel. Quantum metrology uses methods from quantum coherent control to address this difficulty. As an example, entangled states that are invariant under certain noise mechanisms have been engineered with trapped ions and have demonstrated long coherence times^{7–9}. Other entangled states have been similarly engineered to enhance the measurement sensitivity of trapped ions^{10,11}. Whether or not the measurement signal-to-noise ratio improves depends on the commutativity of the noise and signal operators as well as on the noise bandwidth^{12–14}.

A different approach to noise reduction is based on spectrally separating a quantum system from its noise environment. Such time-dynamical noise decoupling has been demonstrated using trapped-ion quantum bits, among other systems, and has been optimized to match different noise profiles^{15,16}. In fact, it was shown that the decoherence rate of these modulated systems can be used to extract information about their noise spectrum^{16–18}. A natural extension to spectral characterization is the measurement of oscillating signals. Dynamical manipulation can therefore be used to decouple a quantum probe from noise while enhancing its sensitivity to alternating signals.

In the past few years, dynamical decoupling methods have been used to improve on the signal-to-noise ratio of a.c. magnetometry

using nitrogen-vacancy centres^{19–22}. Indeed, significant enhancement of sensitivity was achieved using a few tens of modulation pulses²². However, owing to the particular decoherence mechanism in nitrogen-vacancy centres, their best reported magnetic field measurement sensitivity, of $4 \text{ nT Hz}^{-1/2}$, was achieved using a single echo pulse³.

In this work, we show that a quantum probe, time evolving under non-commuting noise, signal and modulation operators, is equivalent to a lock-in amplifier. We take full advantage of the quantum lock-in method, with up to 650 modulation pulses, using a single trapped $^{88}\text{Sr}^+$ ion. The lock-in method provides a 30-fold improvement in frequency-shift measurement sensitivity. We demonstrate a record sensitivity for a single-spin detector^{2,3}, of $15 \text{ pT Hz}^{-1/2}$ ($0.42 \text{ Hz Hz}^{-1/2}$), reaching a measurement uncertainty of less than 10 mHz (350 fT) after 3,720 s of averaging.

Classical lock-in amplifiers are detectors that can extract a signal with a known carrier frequency from an extremely noisy environment. Schematically, if noise, $N(t)$, adds to a physical observable, S_0 , oscillating at a frequency f_m , the total signal measured by the detector is $M(t) = m_0[S_0\cos(2\pi f_m t + \varphi) + N(t)]$. Here m_0 sets the detector measurement units and φ is a constant phase. A signal proportional to S_0 is obtained by a mix-down process: $M(t)$ is multiplied by either $\sin(2\pi f_m t)$ or $\cos(2\pi f_m t)$ and the two results are integrated over an integration window, T :

$$I_{\text{lock-in}} = \frac{1}{T} \int_0^T dt M(t) \cos(2\pi f_m t) \quad (1)$$

$$Q_{\text{lock-in}} = \frac{1}{T} \int_0^T dt M(t) \sin(2\pi f_m t)$$

The signal S_0 is proportional to $(I_{\text{lock-in}}^2 + Q_{\text{lock-in}}^2)^{1/2}$. The constant phase φ can be extracted from $\tan(\varphi) = -Q_{\text{lock-in}}/I_{\text{lock-in}}$. Noise spectral components with frequencies far from f_m will be averaged out in the integration. Therefore, by choosing f_m outside the noise bandwidth, the measurement signal-to-noise ratio can be significantly improved.

The main obstacle in realizing quantum lock-in dynamics is finding a quantum analogue to signal multiplication, which is essential for the mix-down process. In a classical apparatus this is achieved using a nonlinear device with an output that is proportional to the instantaneous product of its inputs. Nonlinear dynamics of the wavefunction cannot be introduced directly, owing to the linearity of Schrödinger's equation. Nevertheless, wavefunction dynamics will be proportional to a product of Hamiltonian terms if the total Hamiltonian does not commute with itself at different times. Operator non-commutativity therefore has an important role in the quantum mix-down process.

To show this in more detail, we turn to the case of a two-level quantum probe, with states $|\uparrow\rangle$ and $|\downarrow\rangle$. We assume that the probe is coupled both to a signal, $S(t)$, and noise, $N(t)$, by $H_{\text{int}} = M(t)\hat{\sigma}_z/2$, where $M(t) = S(t) + N(t)$ and $\hat{\sigma}_x$, $\hat{\sigma}_y$ and $\hat{\sigma}_z$ are the Pauli operators. For a lock-in measurement, $S(t)$ is modulated: $S(t) = S_0\cos(2\pi f_m t + \varphi)$. The probe is initialized to $|\psi_0\rangle = (|\uparrow\rangle + |\downarrow\rangle)/\sqrt{2}$. In a Bloch sphere picture, this state is represented by a vector along the x axis. Under H_{int} , the superposition phase (the angle between the Bloch vector and the x axis) is oscillating back and forth as a result of the signal and is randomly varying owing to the effect of noise. To implement a lock-in

¹Department of Physics of Complex Systems, Weizmann Institute of Science, PO Box 26, Rehovot 76100, Israel.

measurement, we mix the probe phase with an oscillating signal by adding to H_{int} an oscillating term that does not commute with $\hat{\sigma}_z$: $H = (M(t)\hat{\sigma}_z + \Omega(t)\hat{\sigma}_y)/2$. If $\Omega(t)$ is periodic and synchronized with $S(t)$, then the phase accumulated owing to $S(t)$ coherently adds up whereas the random phase accumulated owing to $N(t)$ is averaged away. The probe superposition is characterized by the probability of finding the probe in the $|\uparrow\rangle$ state, P_\uparrow , and the superposition relative phase, $\phi_{\text{lock-in}}$. By measuring both at time T , we extract the quantum lock-in signal:

$$\phi_{\text{lock-in}} = \frac{1}{\hbar} \int_0^T dt M(t) \cos\left(\frac{1}{\hbar} \int_0^t dt' \Omega(t')\right) \quad (2)$$

$$1 - 2P_\uparrow = \frac{1}{\hbar} \int_0^T dt M(t) \sin\left(\frac{1}{\hbar} \int_0^t dt' \Omega(t')\right)$$

Equation (2), where \hbar is Planck's constant divided by 2π , resembles the classical lock-in output in equation (1). Specifically, for a constant $\Omega(t) \equiv \Omega_0$, the lock-in outputs $\phi_{\text{lock-in}}$ and $1 - 2P_\uparrow$ faithfully represent the two signal quadratures. Here, instead of reading out a classical parameter, the quantum lock-in read-out requires repetitive quantum projection measurements. We note that the two signal components can be interchanged through single quantum bit rotations. A full derivation and discussion of equation (2) can be found in Supplementary Information.

In our experiment, we use the two spin states of the electronic ground level of a single $^{88}\text{Sr}^+$ ion, $|\uparrow\rangle = |5s_{1/2}, J = 1/2, M_J = 1/2\rangle$ and $|\downarrow\rangle = |5s_{1/2}, J = 1/2, M_J = -1/2\rangle$, as a two-level quantum probe (Fig. 1a). Here J is the total electron angular momentum quantum number and M_J is its projection along the magnetic field axis. Set-up details can be found in Supplementary Information. An energy difference of 5.72 MHz between the probe states is determined by an external d.c. magnetic field. We are able to perform all possible spin rotations by pulsing a resonant radio-frequency magnetic field and tuning the pulse duration and the radio-frequency field phase, ϕ_{rf} . State initialization and measurement are performed by optical pumping and state-selective fluorescence, respectively. Because the probe states are first-order sensitive to magnetic fields, the main noise mechanism is magnetic field noise, with dominant spectral contributions at the 50-Hz line and its harmonics. Examples for signals that we can measure are modulated magnetic or light fields, respectively measured through their resulting Zeeman or light shifts.

The lock-in sequence is depicted in Fig. 1b. Following optical pumping, a $\pi/2$ rotation initializes the ion probe to $|\psi_0\rangle = (|\uparrow\rangle + |\downarrow\rangle)/\sqrt{2}$. To modulate the ion probe, we apply a train of N π pulses, equally spaced τ_{arm} apart. Here, ideally $\Omega(t) = \sum_{n=1}^N \delta(t - n\tau)\pi$, where $\delta(t)$ is the Dirac delta function. Therefore, the cosine term in equation (2) is a square waveform with a period of $2\tau_{\text{arm}}$ and the sine term vanishes. Consequently, a measured signal has to be modulated at $f_m = 1/2\tau_{\text{arm}}$ and in phase with the ion modulation, that is, $\varphi = 0$. Here $\phi_{\text{lock-in}}$ is proportional to the signal magnitude, S_0 . To measure the probe phase, we complete the sequence with an additional $\pi/2$ rotation, with a relative ϕ_{rf} phase with respect to the initial $\pi/2$ pulse. We then detect the probability of the ion being in the $|\uparrow\rangle$ state, $P_\uparrow = 1/2 + (A/2)\cos(\phi_{\text{rf}} - \phi_{\text{lock-in}})$. By scanning ϕ_{rf} , we are able to retrieve both $\phi_{\text{lock-in}}$ and the cosine fringe contrast, A , using a fitting procedure.

Ideally $A = 1$. In practice, noise processes decrease A . As seen from equation (2), even in the absence of any signal, $N(t)$ will contribute a lock-in phase of

$$\phi_N = \frac{1}{\hbar} \int_0^T dt N(t) \cos\left(\frac{1}{\hbar} \int_0^t dt' \Omega(t')\right)$$

The cosine fringe is therefore reduced in the process of averaging: $A = \langle \cos(\phi_N) \rangle$, where angle brackets denote an average over different noise realizations.

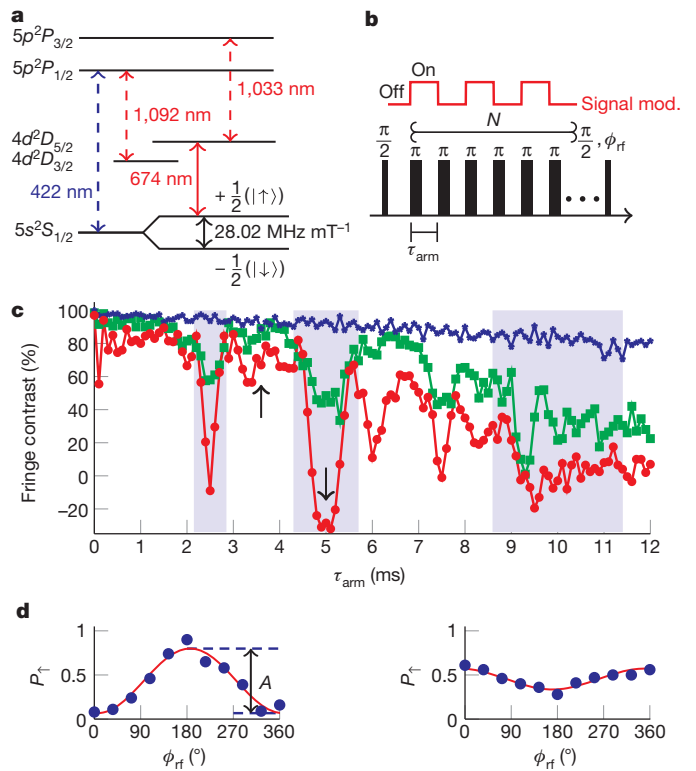


Figure 1 | Measurement scheme. **a**, Level diagram of a $^{88}\text{Sr}^+$ ion. The probe spin states are $|\uparrow\rangle = |5s_{1/2}, J = 1/2, M_J = +1/2\rangle$ and $|\downarrow\rangle = |5s_{1/2}, J = 1/2, M_J = -1/2\rangle$. An external magnetic field splits the two levels by a frequency of $f_0 = 5.72$ MHz. Spin rotations are performed using an oscillating magnetic field. Initialization to $|\uparrow\rangle$ is done by optical pumping. Spin detection is performed by shelving the $|\uparrow\rangle$ state to the metastable level $|D\rangle \equiv |4d_{5/2}, J = 5/2, M_J = +3/2\rangle$, with a narrow-linewidth (<100 -Hz), 674-nm laser, followed by state-selective fluorescence at 422 nm. The 1,092-nm and 1,033-nm lasers are used as repump lasers. **b**, The quantum lock-in measurement pulse scheme. The ion is initialized to $(|\uparrow\rangle + |\downarrow\rangle)/\sqrt{2}$ by a $\pi/2$ pulse. While the measured signal is modulated, the superposition is also modulated, in phase with the signal, by a train of N π pulses, τ_{arm} apart. The total relative phase, $\phi_{\text{lock-in}}$, of the ion superposition, $(|\downarrow\rangle + e^{i\phi_{\text{lock-in}}}|\uparrow\rangle)/\sqrt{2}$, accumulated during the lock-in sequence is measured by scanning the phase of a final $\pi/2$ pulse, ϕ_{rf} , followed by spin detection and a fit of the data to $P_\uparrow = 1/2 + (A/2)\cos(\phi_{\text{rf}} - \phi_{\text{lock-in}})$. **c**, Fringe contrast, A , versus half lock-in modulation period, τ_{arm} , in the absence of any modulated signal. Data corresponding to $N = 1, 9$ and 17 π pulses are shown using blue stars, green rectangles and red circles, respectively. We observe contrast drops as τ_{arm} approaches 2.5, 5 and 10 ms corresponding to magnetic field noise components at 200, 100 and 50 Hz, respectively. **d**, Probability of finding the ion in the $|\uparrow\rangle$ state versus ϕ_{rf} . Fringe plots for $\tau_{\text{arm}} = 3.6$ ms (left) and 5 ms (right), made with lock-in sequences of $N = 17$ π pulses, are shown. The solid line is a best fit to $P_\uparrow = 1/2 + (A/2)\cos(\phi_{\text{rf}})$. The fitted A values are shown in **c** at the locations indicated by the two black arrows. The inverted sign of the second fringe can be understood in terms of equation (4).

The reduction in the fringe contrast has significant implications for the lock-in measurement sensitivity. The lock-in signal, $\phi_{\text{lock-in}}$, is proportional to the energy shift experienced by the probe and can therefore be expressed in terms of frequency or magnetic field. Equation (2) implies that the conversion factor depends on the actual modulation type being used. This is discussed in Supplementary Information, where we also show that the optimal frequency-shift measurement sensitivity, s , is

$$s = \frac{1}{2\pi} \sqrt{\frac{4 - A^2}{2A^2T}} \text{ Hz Hz}^{-1/2} \quad (3)$$

Here $T \equiv (N + 1)\tau_{\text{arm}}$ is the total sequence duration and N is the number of π pulses. The standard quantum limit on the sensitivity is reached when $A = 1$. To optimize sensitivity, the lock-in modulation frequency

and the sequence duration should be chosen so as to minimize the spectral overlap of noise and modulation and therefore maximize A .

We initially quantify the noise floor of our lock-in detector at different modulation frequencies, f_m , and lock-in sequence durations, T , in the absence of any modulated signal. To begin with we perform this measurement at low lock-in modulation frequencies, which are comparable to typical magnetic noise frequencies in our laboratory. We measure A for values of the π -pulse interspacing, τ_{arm} , ranging from 0 to 12 ms, and for $N = 1$ –17 π pulses per lock-in sequence. Both the lock-in sensitivity and the spectral resolution increase as N increases. As shown in Fig. 1c, dips in the fringe contrast emerge as we increase N . These dips, marked by shading, correspond to a.c. magnetic field noise components at frequencies of 200, 100 and 50 Hz, respectively. Figure 1d shows two phase scans for an $N = 17$ lock-in sequence. One scan is at $\tau_{\text{arm}} = 3.6$ ms, where no noise is present, and the other is at $\tau_{\text{arm}} = 5$ ms, where the lock-in modulation has the same period as the 100-Hz noise component.

To use the lock-in method to quantify the magnetic noise spectrum, we assume that it is composed mainly of discrete frequency components, $f_n = \omega_n/2\pi$, with corresponding amplitudes B_n . With this assumption we can calculate

$$A(N, \tau_{\text{arm}}) = \prod_n J_0 \left(\frac{4g\mu_B B_n}{\hbar} \frac{\sin^2 \left(\frac{\omega_n \tau_{\text{arm}}}{2} \right)}{\omega_n} \right) \frac{\sin(N\omega_n \tau_{\text{arm}})}{\sin(\omega_n \tau_{\text{arm}})} \quad (4)$$

Here J_0 is the zeroth Bessel function of the first kind, g is the Landé g -factor and μ_B is the Bohr magneton. We note that A can have negative values, as demonstrated by the inverted sign of the second scan of Fig. 1d. In Fig. 2a, we show A (filled circles) for a lock-in sequence with $N = 17$ and a best fit to equation (4) (solid line). Here we assume four discrete magnetic noise spectral components with respective frequencies of 50, 100 and 150 Hz and f_{slow} , the last a slowly varying field. The noise amplitudes are taken as fit parameters, yielding $B_{50 \text{ Hz}} = 540(3)$ pT, $B_{100 \text{ Hz}} = 390(5)$ pT, $B_{150 \text{ Hz}} = 260(4)$ pT and $g\mu_B B_{\text{slow}}/h = 37(4)$ Hz². The relatively low magnetic field amplitudes are due to an active magnetic field noise cancellation system. A

detailed report of the findings described in this paragraph is under way (S.K., N.A., Y.G. and R.O., manuscript in preparation).

Observing that noise amplitudes at frequencies of more than 200 Hz are negligible, we turn to higher modulation frequencies, in search of the greatest attainable probe coherence time. We modulate the ion probe at $f_m = 312.5$ Hz ($\tau_{\text{arm}} = 1.6$ ms). Figure 2b shows the fringe contrast A versus N , up to $N = 650$. Here, owing to the large number of π pulses, ϕ_{rf} alternates by $\pi/2$ between consecutive pulses, to prevent rotation errors from coherently accumulating. A fit to an exponential decrease in fringe contrast yields a probe coherence time of 1.4(2) s. This is three orders of magnitude longer than the coherence time in the absence of lock-in modulation, measured using Ramsey spectroscopy.

From the data presented so far, we can report our probe's best sensitivity. We calculate the lock-in sensitivity versus T , the total lock-in sequence duration, from the fringe contrast, A , using equation (3). Figure 2c shows the lock-in sensitivity in the low modulation frequency range. A minimum of $0.78 \text{ Hz Hz}^{-1/2}$ ($28 \text{ pT Hz}^{-1/2}$) is observed at $T = 120$ ms, between noise components. Figure 2d shows the lock-in sensitivity versus T at $f_m = 312.5$ Hz. Here a best sensitivity of $0.42(3) \text{ Hz Hz}^{-1/2}$ ($15(1) \text{ pT Hz}^{-1/2}$) is observed at the minimum of the fit, with a lock-in sequence duration of $T = 624$ ms. This is, to our knowledge^{2,3}, the best magnetic field sensitivity reported so far using a single-spin (or pseudo-spin) detector. In both cases, the measured sensitivity differs from the standard quantum limit, shown by the dashed line, by a factor of less than 1.5.

We next demonstrate the lock-in detection of a small signal and experimentally verify equation (2). To this end, we measure the light shift of a narrow-linewidth (<100 -Hz) laser nearly resonant with the $|\uparrow\rangle \rightarrow |\text{D}\rangle = |4d_{5/2}, J = 5/2, M_J = 3/2\rangle$ quadrupole transition at 674 nm. The laser amplitude is switched on and off at a rate $f_L = 500$ Hz. With this scheme, both the lock-in and the laser are square-wave modulated. We apply a lock-in sequence of $N = 99$ π pulses and scan the lock-in modulation frequency. Here the 674-nm laser is detuned by $\Delta = -17$ kHz from resonance (red detuned). A laser Rabi frequency of $2\pi \times 840$ Hz is independently measured by

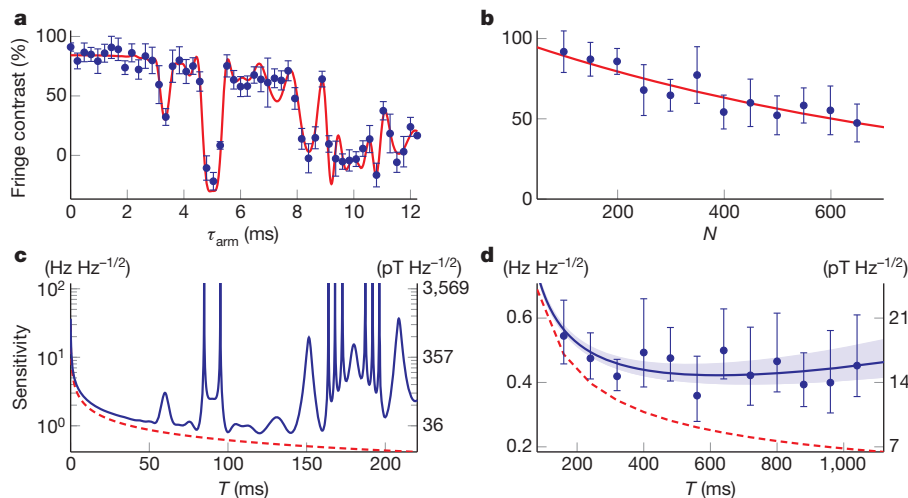


Figure 2 | Sensitivity of the quantum lock-in measurement. **a**, Fringe contrast, A , versus τ_{arm} for $N = 17$ π pulses, in the absence of any modulated signal. Each point is the fitted contrast of a corresponding measured fringe as in Fig. 1; error bars are 95% confidence intervals. The data are used to extract the magnetic noise spectrum. The solid red line is a best fit to equation (4) with four fit parameters: the field amplitudes $B_{50 \text{ Hz}} = 540(3)$ pT, $B_{100 \text{ Hz}} = 390(5)$ pT and $B_{150 \text{ Hz}} = 260(4)$ pT and a slowly varying field $g\mu_B B_{\text{slow}}/h = 37(4)$ Hz². **b**, Fringe contrast, A , versus number of π pulses, N , at a lock-in modulation frequency of $f_m = 312.5$ Hz; error bars are 95% confidence intervals. The red line is an exponential decay fit to the data yielding a $1/e$ coherence decay time of 1.4(2) s. **c**, Lock-in sensitivity (solid blue line) versus the lock-in sequence duration, T , calculated from **a** using equation (3). The dashed red line is the

standard quantum limit on sensitivity (achieved when $A = 1$). A best sensitivity of $0.78 \text{ Hz Hz}^{-1/2}$ ($28 \text{ pT Hz}^{-1/2}$) is observed at $T = 120$ ms. This sensitivity is only a factor of 1.5 greater than the standard quantum limit. The sensitivity diverges whenever A (shown in **a**) crosses zero. **d**, Exponential decay fit (solid blue curve) shown in **b**, translated to sensitivity using equation (3), as in **c**. The shaded region is a 95% confidence interval for the curve. The dashed red line shows the standard quantum limit on the lock-in sensitivity. The solid blue circles are calculated sensitivities of the measured fringe contrast points in **b**, with 95% confidence intervals. A best sensitivity of $0.42(3) \text{ Hz Hz}^{-1/2}$ ($15(1) \text{ pT Hz}^{-1/2}$) is obtained at the minimum of the solid blue curve ($T = 624$ ms). A similar value of $0.4(1) \text{ Hz Hz}^{-1/2}$ ($13(3) \text{ pT Hz}^{-1/2}$) is observed at $T = 560$ ms.

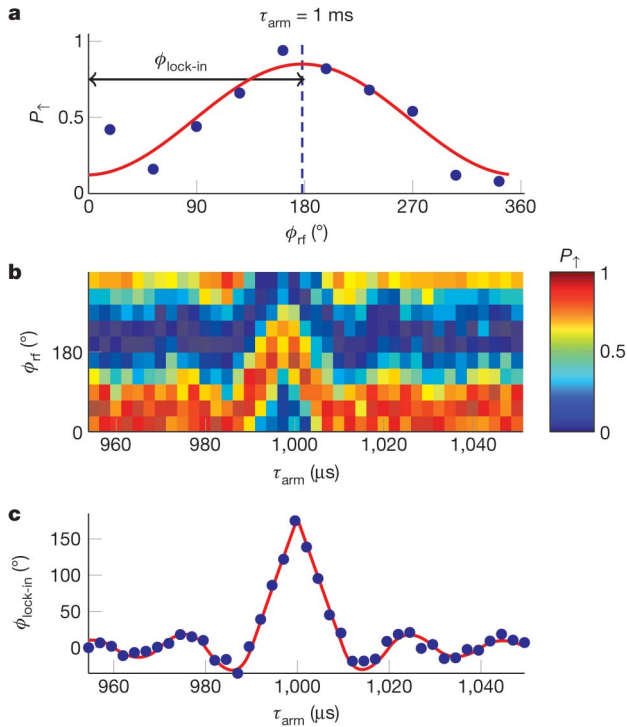


Figure 3 | Lock-in measurement of a small signal. The light shift of the $|\uparrow\rangle$ state induced by the 674-nm laser is measured. The laser is detuned by $\Delta_{674\text{ nm}} = -17\text{ kHz}$ from the $|\uparrow\rangle \rightarrow |D\rangle$ quadrupole transition, and is amplitude-modulated by a square wave of frequency $f_L = 500\text{ Hz}$. The lock-in scheme has $N = 99\pi$ pulses and the lock-in frequency, $f_m = 1/2\tau_{\text{arm}}$, is varied. **a**, Lock-in fringe scan, P_{\uparrow} , versus ϕ_{rf} , at a lock-in period of $2\tau_{\text{arm}} = 2\text{ ms}$. The red solid line is a best fit to $P_{\uparrow} = 1/2 + (A/2)\cos(\phi_{\text{rf}} - \phi_{\text{lock-in}})$. A clear phase shift of $\phi_{\text{lock-in}} = 0.99\pi$ is observed, with a fringe contrast of $A = 72\%$. **b**, The columns are lock-in fringe scans, similar to that in **a**, for various values of τ_{arm} . The lock-in signal, $\phi_{\text{lock-in}}$, is seen to increase as the lock-in modulation frequency, f_m , approaches the laser modulation frequency, $f_L = 500\text{ Hz}$. **c**, Lock-in signal, $\phi_{\text{lock-in}}$, versus τ_{arm} , extracted from **b** as explained in **a**. A light shift of $9.7(4)\text{ Hz}$ is measured (with 95% confidence). The solid red line is calculated using equation (2) without any fit parameters.

an on-resonance Rabi nutation curve. Figure 3a shows a fringe scan at a lock-in modulation frequency of $f_m = 500\text{ Hz}$. The solid line is a best fit to $P_{\uparrow} = 1/2 + (A/2)\cos(\phi_{\text{rf}} - \phi_{\text{lock-in}})$, with A and $\phi_{\text{lock-in}}$ as fit parameters. A clear phase shift of 0.99π is observed. The columns in Fig. 3b are fringe scans similar to that in Fig. 3a, made at different lock-in modulation frequencies. As seen, the lock-in signal is maximal when the modulation frequency approaches 500 Hz ($\tau_{\text{arm}} = 1,000\text{ }\mu\text{s}$), that is, the modulation rate of the laser. Figure 3c shows that the prediction of equation (2) (solid line, calculated without any fit parameters) is in good agreement with measured values (filled circles) of $\phi_{\text{lock-in}}$ as a function of the lock-in modulation rate. A light shift of $9.7(4)\text{ Hz}$ is measured; the theoretically predicted value is $9.9(4)\text{ Hz}$.

Any measurement uncertainty is ultimately limited, at long integration times, by slow systematic drifts. The optimal averaging time can be found by performing an Allan deviation analysis²³. We obtain a minimal measurement uncertainty of $8(2)\text{ mHz}$ ($290(70)\text{ fT}$) after 3,720 s of averaging. We perform the same analysis for a light shift measurement, obtaining $0.12(2)\text{ Hz}$ after 1,320 s. This uncertainty is most probably limited by slow frequency drifts of the 674-nm laser (see Supplementary Information for Allen plots and more details). The magnetic field generated by the valence electron spin of a single $^{88}\text{Sr}^+$ ion will cause a level shift of 52 mHz in a probe ion co-trapped one micrometre away, the measurement of which could be within our experimental reach.

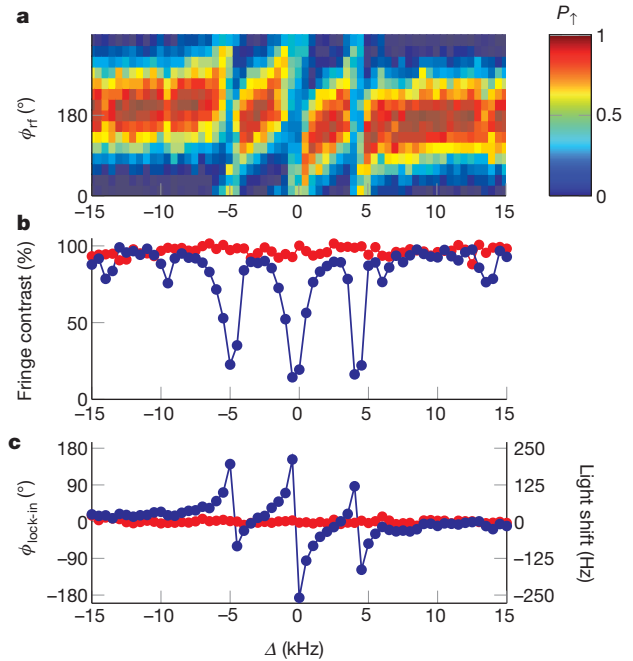


Figure 4 | Light shift spectroscopy. Light shift of $|\uparrow\rangle$ induced by the 674-nm laser, as a function of laser frequency detuning, Δ . At each Δ value, a lock-in sequence of $N = 39\pi$ pulses with a lock-in period of $2\tau_{\text{arm}} = 200\text{ }\mu\text{s}$ is applied while the 674-nm laser is amplitude-modulated at the same frequency. **a**, Every column is a lock-in fringe scan for a particular value of Δ . For each column, the lock-in signal, $\phi_{\text{lock-in}}$, is the shift of the fringe minimum from zero. **b**, Fringe contrast, A (blue filled circles), versus Δ . Red filled circles show A in the absence of laser light. We observe a reduction in contrast due to shelving of the $|\uparrow\rangle$ state to the metastable level $|D\rangle$ whenever the laser approaches resonance. **c**, Lock-in signal, $\phi_{\text{lock-in}}$ (blue filled circles), versus Δ . The red filled circles show $\phi_{\text{lock-in}}$ in the absence of laser light. Light shifts are seen to have dispersive resonance. Both **b** and **c** show two sidebands, separated by 5 kHz from the transition carrier, generated by the fast amplitude modulation of the laser.

Finally, we show how the lock-in method can be used to perform light shift spectroscopy. We probe the $|\uparrow\rangle \rightarrow |D\rangle$ transition. Figure 4a shows lock-in phase scans (columns) for different laser detunings. Figure 4b and Fig. 4c show the fringe contrast, A , and the lock-in signal, $\phi_{\text{lock-in}}$, respectively. Population transfer to level $|D\rangle$ results in a reduction in A whenever the laser is close to resonance. The measured light shift is seen to be dispersive around resonance. The three resonances, a carrier and two sidebands, are due to the fast amplitude modulation of the laser, reminiscent of the Pound–Drever–Hall signal of a laser scanning across an optical cavity resonance²⁴. Such a dispersive signal can be used to lock a narrow-linewidth laser to an atomic clock transition.

The results presented here demonstrate the potency of the quantum lock-in measurement technique, which is readily available for any quantum probe. Specifically, with single trapped-ions the lock-in technique allows high-precision frequency-shift measurements with nanometre-scale spatial resolution (in our set-up, the ion wavefunction extent is 9 nm with ground-state cooling). In addition to the detection of a single electronic spin mentioned above, this would be useful to probe spin-dependent interactions of an ion submerged in a quantum degenerate gas^{25,26}. Finally, the quantum lock-in technique can be useful for precision measurements and frequency metrology. As an example, it can be used to measure the very small frequency shifts required for the observation of parity non-conservation in a single trapped ion⁴. Another example is the characterization of systematic errors, such as the quadrupole shift, in ion-based atomic clocks²⁷. As a final example, the technique can be used to characterize the noise spectrum of narrow-linewidth lasers with respect to an atomic transition.

Received 24 January; accepted 22 March 2011.

1. Giovannetti, V., Lloyd, S. & Maccone, L. Quantum-enhanced measurements: beating the standard quantum limit. *Science* **306**, 1330–1336 (2004).
2. Degen, C. Nanoscale magnetometry microscopy with single spins. *Nature Nanotechnol.* **3**, 643–644 (2008).
3. Balasubramanian, G. *et al.* Ultralong spin coherence time in isotopically engineered diamond. *Nature Mater.* **8**, 383–387 (2009).
4. Fortson, N. Possibility of measuring parity nonconservation with a single trapped atomic ion. *Phys. Rev. Lett.* **70**, 2383–2386 (1993).
5. Rosenband, T. *et al.* Frequency ratio of Al^+ and Hg^+ single-ion optical clocks; metrology at the 17th decimal place. *Science* **319**, 1808–1812 (2008).
6. Knünz, S. *et al.* Injection locking of a trapped-ion phonon laser. *Phys. Rev. Lett.* **105**, 013004 (2010).
7. Kielpinski, D. *et al.* A decoherence-free quantum memory using trapped ions. *Science* **291**, 1013–1015 (2001).
8. Roos, C. F. *et al.* Bell states of atoms with ultralong lifetimes and their tomographic state analysis. *Phys. Rev. Lett.* **92**, 220402 (2004).
9. Langer, C. *et al.* Long-lived qubit memory using atomic ions. *Phys. Rev. Lett.* **95**, 060502 (2005).
10. Leibfried, D. *et al.* Toward Heisenberg-limited spectroscopy with multiparticle entangled states. *Science* **304**, 1476–1478 (2004).
11. Roos, C. F., Chwalla, M., Kim, K., Riebe, M. & Blatt, R. 'Designer atoms' for quantum metrology. *Nature* **443**, 316–319 (2006).
12. Huelga, S. F. *et al.* Improvement of frequency standards with quantum entanglement. *Phys. Rev. Lett.* **79**, 3865–3868 (1997).
13. André, A., Sørensen, A. S. & Lukin, M. D. Stability of atomic clocks based on entangled atoms. *Phys. Rev. Lett.* **92**, 230801 (2004).
14. Wineland, D. *et al.* Experimental issues in coherent quantum-state manipulation of trapped atomic ions. *J. Res. Natl Inst. Stand. Technol.* **103**, 259–328 (1998).
15. Uys, H., Biercuk, M. J. & Bollinger, J. J. Optimized noise filtration through dynamical decoupling. *Phys. Rev. Lett.* **103**, 040501 (2009).
16. Biercuk, M. J. *et al.* Optimized dynamical decoupling in a model quantum memory. *Nature* **458**, 996–1000 (2009).
17. Gordon, G., Erez, N. & Kurizki, G. Universal dynamical decoherence control of noisy single- and multi-qubit systems. *J. Phys. At. Mol. Opt. Phys.* **40**, S75–S93 (2007).
18. Sagi, Y., Almog, I. & Davidson, N. Process tomography of dynamical decoupling in a dense cold atomic ensemble. *Phys. Rev. Lett.* **105**, 053201 (2010).
19. Maze, J. R. *et al.* Nanoscale magnetic sensing with an individual electronic spin in diamond. *Nature* **455**, 644–648 (2008).
20. Hall, L. T., Hill, C. D., Cole, J. H. & Hollenberg, L. C. L. Ultrasensitive diamond magnetometry using optimal dynamic decoupling. *Phys. Rev. B* **82**, 045208 (2010).
21. Naydenov, B. *et al.* Dynamical decoupling of a single-electron spin at room temperature. *Phys. Rev. B* **83**, 081201(R) (2011).
22. de Lange, G., Ristè, D., Dobrovitski, V. V. & Hanson, R. Single-spin magnetometry with multipulse dynamical decoupling sequences. *Phys. Rev. Lett.* **106**, 080802 (2011).
23. Riley, W. J. *Handbook of Frequency Stability Analysis*. NIST Spec. Publ. 1065 (US Department of Commerce, National Institute of Standards and Technology, 2008).
24. Drever, R. W. P. *et al.* Laser phase and frequency stabilization using an optical resonator. *Appl. Phys. B* **31**, 97–105 (1983).
25. Zipkes, C., Palzer, S., Sias, C. & Koehl, M. A trapped single ion inside a Bose–Einstein condensate. *Nature* **464**, 388–391 (2010).
26. Schmid, S., Härter, A. & Denschlag, J. H. Dynamics of a cold trapped ion in a Bose–Einstein condensate. *Phys. Rev. Lett.* **105**, 133202 (2010).
27. Oskay, W. H., Itano, W. M. & Bergquist, J. C. Measurement of the $^{199}\text{Hg}^+ 5d^96s^2 \ ^2D_{5/2}$ electric quadrupole moment and a constraint on the quadrupole shift. *Phys. Rev. Lett.* **94**, 163001 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Bensky, G. Gordon and G. Kurizki for discussions. We acknowledge the support by the ISF Morasha program, the Crown Photonics Center and the Minerva Foundation.

Author Contributions All authors participated in the building of the experimental apparatus. S.K. led the data-taking effort, with help from N.A. Data analysis and development of the analytic theory were performed by S.K. S.K. and R.O. wrote the manuscript. R.O. designed the experiment and supervised the work. All authors participated in discussions, contributed ideas along the way and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.K. (shlomi.kotler@weizmann.ac.il).

Convergence of electronic bands for high performance bulk thermoelectrics

Yanzhong Pei¹, Xiaoya Shi², Aaron LaLonde¹, Heng Wang¹, Lidong Chen² & G. Jeffrey Snyder¹

Thermoelectric generators, which directly convert heat into electricity, have long been relegated to use in space-based or other niche applications, but are now being actively considered for a variety of practical waste heat recovery systems—such as the conversion of car exhaust heat into electricity. Although these devices can be very reliable and compact, the thermoelectric materials themselves are relatively inefficient: to facilitate widespread application, it will be desirable to identify or develop materials that have an intensive thermoelectric materials figure of merit, zT , above 1.5 (ref. 1). Many different concepts have been used in the search for new materials with high thermoelectric efficiency, such as the use of nanostructuring to reduce phonon thermal conductivity^{2–4}, which has led to the investigation of a variety of complex material systems⁵. In this vein, it is well known^{6,7} that a high valley degeneracy (typically ≤ 6 for known thermoelectrics) in the electronic bands is conducive to high zT , and this in turn has stimulated attempts to engineer such degeneracy by adopting low-dimensional nanostructures^{8–10}. Here we demonstrate that it is possible to direct the convergence of many valleys in a bulk material by tuning the doping and composition. By this route, we achieve a convergence of at least 12 valleys in doped $\text{PbTe}_{1-x}\text{Se}_x$ alloys, leading to an extraordinary zT value of 1.8 at about 850 kelvin. Band engineering to converge the valence (or conduction) bands to achieve high valley degeneracy should be a general strategy in the search for and improvement of bulk thermoelectric materials, because it simultaneously leads to a high Seebeck coefficient and high electrical conductivity.

A high thermoelectric figure of merit, ZT , for a high-efficiency thermoelectric generator requires the constituent n-type and p-type materials each to have a high average thermoelectric materials figure of merit, $zT = S^2\sigma T/(\kappa_E + \kappa_L)$, where T , S , σ , κ_E and κ_L are the temperature, Seebeck coefficient, electrical conductivity, and the electronic and lattice components of the thermal conductivity, respectively. To date, commercial products for thermoelectric power generation utilize only PbTe- or Bi_2Te_3 -based materials with peak zT of less than unity⁵.

Recent efforts to raise the zT value of PbTe have focused on nanostructured composites, such as $\text{Na}_{1-x}\text{Pb}_m\text{Sb}_y\text{Te}_{m+2}$ (ref. 3), where the aim is to reduce κ_L and thus to enhance zT ; indeed $zT > 1$ has been obtained in many instances². Such materials have κ_L close to the amorphous limit^{2,4}, lending greater potential to the increasing of zT by the enhancement of the electronic component ($S^2\sigma$). Seebeck coefficient enhancement through density of states modification^{8,11} is a promising route, but this approach risks the reduction of carrier mobility.

The optimal electronic performance of a thermoelectric semiconductor depends primarily on the weighted mobility^{6,7,12}, $\mu(m^*/m_e)^{3/2}$; here m^* is the density-of-states effective mass, μ is the mobility of carriers, and m_e is the electron mass. However, μ is low for bands with heavy mass m_b^* (the band-mass of a single valley, or mass of a single pocket of Fermi surface related to $1/(d^2E/dk^2)$ of the pocket). In fact, for charge carriers predominantly scattered by acoustic phonons

(as has been found to occur in most good thermoelectric materials), it is expected that $\mu \propto 1/m_b^{*5/2}$ (ref. 7). Therefore, increasing the band-mass should be detrimental to the thermoelectric performance⁷.

In contrast, the convergence of many charge carrying valleys has virtually no detrimental effects. Multiple degenerate valleys (separate pockets of Fermi surface with the same energy) have the effect of producing large m^* without explicitly reducing μ . A valley degeneracy N_v has the effect of increasing m^* by a factor of $N_v^{2/3}$. Specifically, the density-of-states effective mass used to analyse most thermoelectric data is given by $m^* = N_v^{2/3}m_b^*$ (refs 6, 7, 12, 13), where N_v includes orbital degeneracy, and m_b^* is, more specifically, the average (single valley) density-of-states effective mass of the degenerate valleys (including the effect of spin degeneracy but not orbital degeneracy or degeneracy imposed by the symmetry of the Brillouin zone)⁶. The mobility is nominally unaffected by N_v , but there may be some reduction due to intervalley scattering.

It is thus clear that a large valley degeneracy is good for thermoelectric materials^{6,7,12,13}. More generally, bands may be regarded as effectively converged when their energy separation is small (compared with $k_B T$, where k_B is the Boltzmann constant); this leads to an effective increase in N_v , even when the bands are not exactly degenerate. The concept of carrier pocket engineering to produce convergence (high N_v) of symmetrically inequivalent bands has been suggested in the context of manipulating low-dimensional thermoelectric nanostructures^{8–10}. Extending this concept to bulk materials would be most useful for rapid integration into commercial devices.

Convergence of many valleys can occur in high symmetry crystal structures (such as PbTe and $(\text{Bi}, \text{Sb})_2\text{Te}_3$) if the Fermi surface forms isolated pockets at low symmetry points. The widely used thermoelectric material $(\text{Bi}, \text{Sb})_2\text{Te}_3$ has significant valley degeneracy, with $N_v = 6$ in both the conduction and valence bands⁷. The valence band extremum in PbTe occurs at the L point in the Brillouin zone, where

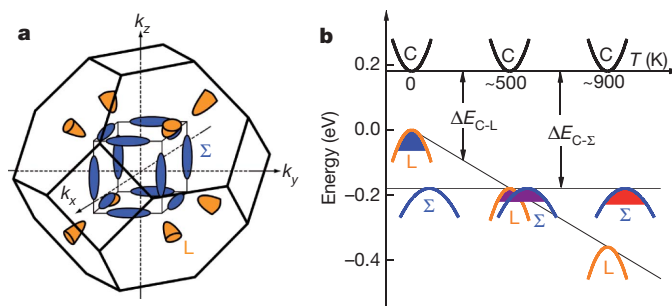


Figure 1 | Valence band structure of $\text{PbTe}_{1-x}\text{Se}_x$. **a**, Brillouin zone showing the low degeneracy hole pockets (orange) centred at the L point, and the high degeneracy hole pockets (blue) along the Σ line. The figure shows 8 half-pockets at the L point so that the full number of valleys, N_v , is 4, while the valley degeneracy of the Σ band is $N_v = 12$. **b**, Relative energy of the valence bands in $\text{PbTe}_{0.85}\text{Se}_{0.15}$. At ~ 500 K the two valence bands converge, resulting in transport contributions from both the L and Σ bands. C, conduction band; L, low degeneracy hole band; Σ , high degeneracy hole band.

¹Materials Science, California Institute of Technology, Pasadena, California 91125, USA. ²CAS Key Laboratory of Materials for Energy Conversion, Shanghai Institute of Ceramics, Chinese Academy of Sciences, Shanghai 200050, China.

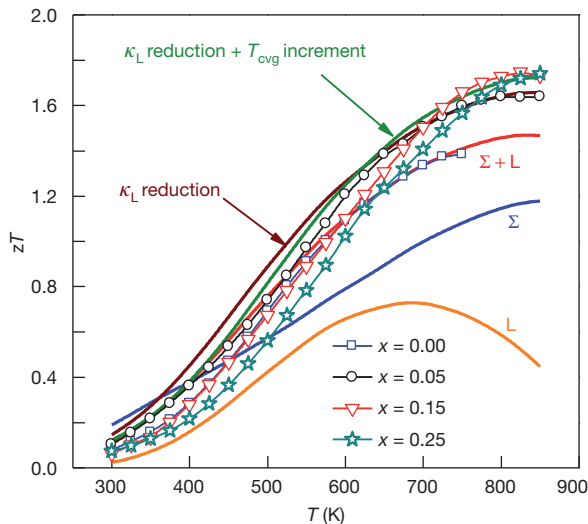


Figure 2 | Temperature dependence of the zT of $p\text{-PbTe}_{1-x}\text{Se}_x$ materials doped with 2 atom % Na. Symbols are experimental data. Curves are calculated results, based on a three-band model with a total hole density of $2.5 \times 10^{20} \text{ cm}^{-3}$. The zT values calculated for both L and Σ bands ($\Sigma + L$) are shown with individual contributions from the L and Σ bands for $x = 0$. The zT for $x = 0.15$ using the predicted κ_L in Fig. 3b is calculated with (κ_L reduction + T_{cvg} increment) and without (κ_L reduction) band structure modification (equation (3)) respectively to compare the individual contributions.

the valley degeneracy is 4 (refs 6,14–16), as in the conduction band of Ge (ref. 17)). However, in addition there is in PbTe a second valence band along the Σ line; this second valence band (the Σ band) has an energy about 0.2 eV below that of the first valence band (the L band), and has a valley degeneracy of 12 (Fig. 1a)^{14,15}.

By producing the convergence of many valleys at the desired temperatures, thermoelectric performance can be greatly enhanced if properly doped. We demonstrate this effect in $\text{PbTe}_{1-x}\text{Se}_x$, where the L and Σ valence bands (Fig. 1b) can be converged, giving an increased valley degeneracy of 16. This exceptionally high degeneracy persists to high temperature, at which the effective degeneracy is at least 12 from the Σ band alone. Combining this effect with the low lattice thermal conductivity of $\text{PbTe}_{1-x}\text{Se}_x$ alloys, we observe a zT value of ~ 1.8 at temperatures above ~ 800 K (Fig. 2).

In a system that contains two valence (or conduction) bands, the total electrical conductivity (σ_{total}) and Seebeck coefficient (S_{total}) can be expressed as:

$$\sigma_{\text{total}} = \sigma_1 + \sigma_2 \quad (1)$$

$$S_{\text{total}} = (\sigma_1 S_1 + \sigma_2 S_2) / \sigma_{\text{total}} \quad (2)$$

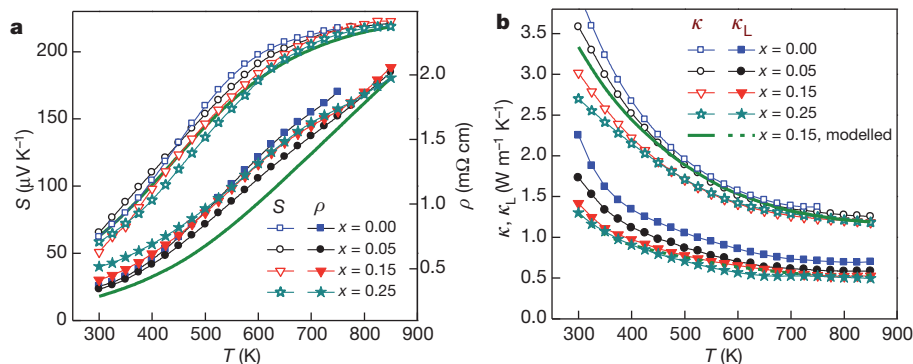


Figure 3 | Thermoelectric transport properties of $\text{PbTe}_{1-x}\text{Se}_x$ alloy doped with 2 atom % Na. a, Temperature dependence of the Seebeck coefficient S and resistivity ρ . b, Total thermal conductivity κ and its lattice component κ_L . The

Here, subscripts 1 and 2 refer to the transport properties of carriers in the individual band. If two bands are present, then the total Seebeck coefficient is a weighted average of the Seebeck coefficients of the individual band; the band with the higher conductivity is more strongly weighted. Because S usually decreases with the number of carriers n , whereas conductivity increases with n ($\sigma = ne\mu$), the total Seebeck coefficient will generally be closer to the smaller S of the two bands. Only when the two band energies are aligned (degenerate), such that the two bands have the same Seebeck coefficient, will S_{total} be maintained while the total conductivity is substantially higher than that of either band alone. In general, this effect will improve thermoelectric performance when the bands are within $\sim 2k_B T$ of each other, owing to the broadening of the Fermi distribution, making band convergence easier to achieve at higher temperatures. Additionally, higher doping concentrations place the Fermi level deeper within the first band, helping to position the Fermi level within $\pm 2k_B T$ of both bands.

The existence of the secondary (Σ) valence band slightly below the principal (L) valence band in PbTe has been confirmed by recent density functional theory calculations¹⁸. A schematic band structure¹⁵ of PbTe is depicted in Fig. 1b. With this two valence band model, the electrical transport, optical spectroscopy and other properties of $p\text{-PbTe}$ can be well understood. Most importantly, the L band moves below the Σ band at $T \gtrsim 450$ K. Temperature (T , in K)-dependent energy offsets (in eV) of the L and Σ bands from the conduction (C) band are given by^{14–16,19}:

$$\Delta E_{C-L} = 0.18 + (4T/10,000) - 0.04x$$

$$\Delta E_{C-\Sigma} = 0.36 + 0.10x \quad (3)$$

where x refers to the subscript x in $\text{PbTe}_{1-x}\text{Se}_x$. Partial substitution of Se for Te ($\text{PbTe}_{1-x}\text{Se}_x$) increases the energy of the L band and reduces the energy of the Σ band (resulting in the terms including x in equation (3) above), according to a linear dependence of band energy versus Se content²⁰. Therefore, alloying with Se will increase the convergence temperature (T_{cvg}) of the L and Σ bands. Na is an effective p-type dopant in PbTe, and can be used to obtain a hole density above 10^{20} cm^{-3} by replacing nominally divalent Pb with monovalent Na (ref. 21). The valence band at the L point with $N_v = 4$ has sufficient mobility to enable a good zT of about 0.8 (curve ‘L’ in Fig. 2 is the contribution from the low degeneracy L band, calculated using the model described in Supplementary Information). The second valence band along the Σ line with a higher $N_v = 12$ has even higher performance (curve ‘ Σ ’ in Fig. 2) at such doping levels.

When both the L and Σ bands are aligned the carriers are redistributed, populating the highly degenerate Σ valleys, which creates a Seebeck coefficient that increases faster than the typical linear temperature dependence (Fig. 3a). With the combined 16 hole pockets

solid green curves show the results of the three-band model and the dashed green curve shows the predicted κ_L according to the Debye–Callaway model for $x = 0.15$ with a hole concentration of $2.5 \times 10^{20} \text{ cm}^{-3}$.

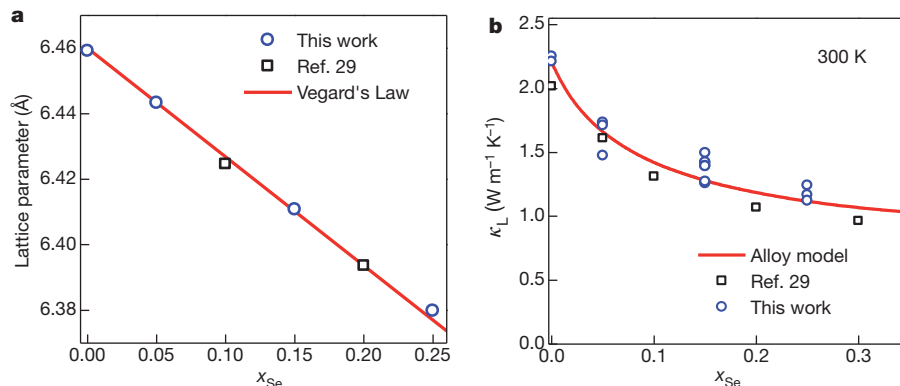


Figure 4 | Composition dependence of lattice parameter and lattice thermal conductivity for $\text{PbTe}_{1-x}\text{Se}_x$ doped with Na, compared with models expected for alloys. Literature results²⁹ are included in this figure. **a**, The 300 K lattice parameter obeys Vegard's law, which is expected for a solid solution.

contributing, a high zT of ~ 1.3 at ~ 700 K (for example, curve ' $\Sigma + \text{L}$ ' in Fig. 2) can be obtained, similar to that seen in PbTe heavily doped with alkali metals (Na or K)^{22,23}.

As mentioned above, alloying with Se increases T_{cvg} , further tuning the thermoelectric properties so as to increase the zT in the temperature range for waste heat recovery (400–900 K). The energy difference between the L and Σ bands, $\Delta E_{\text{L}-\Sigma}$, of the $(\text{PbTe}_{1-x}\text{Se}_x)$ alloy is reduced to $< 2k_{\text{B}}T$ even at high temperatures, making the two bands effectively converged. Alloying with Se has the extra benefit of increasing the bandgap (helpful for higher temperature operation) and also provides lower lattice thermal conductivity due to point defect scattering of phonons. Further increasing the Se content may improve the peak zT but additional Se leads to lower mobility from impurity scattering of electrons, and therefore no significant benefit in average zT is realized, similar to that found in under-doped alloys²⁴.

To confirm that the multi-band effects are indeed responsible for the extraordinary thermoelectric properties, we have developed a detailed three-band model ($\text{C} + \text{L} + \Sigma$). It is important to include the temperature dependence of the bandgap, band offsets and effective masses to fit the data accurately. These parameters have been determined by optical absorption spectroscopy and other temperature dependent transport properties for a wide range of carrier densities^{14–16,19,25,26}. Bands L and C have been found to be non-parabolic and have been described by the Kane model^{14,27}, whereas the high degeneracy hole band Σ has been described as parabolic^{14,28}. It is also assumed that acoustic phonons dominate the electron scattering^{14,27}. This model also gives the Lorenz factor (Supplementary Fig. 3) needed to calculate the electronic contribution to the thermal conductivity. The details of this model are given in Supplementary Information.

The X-ray lattice parameter (Fig. 4a, Supplementary Fig. 1) of our annealed samples follows the simple form of Vegard's law, which predicts a linear change from 6.46 Å for PbTe (ref. 14) to 6.12 Å for PbSe (ref. 14), suggesting the formation of a simple alloy consistent with previous studies^{14,29}. Na is also expected to be homogeneously distributed because of the high dopant effectiveness of monovalent (Na^+) on the Pb site (Hall effect data are given in Supplementary Fig. 3), and because there are no trivalent species present to induce clustering as found in some other similar systems²³, such as $\text{Na}_{1-x}\text{Pb}_m\text{Sb}_y\text{Te}_{m+2}$. Nevertheless, we cannot entirely rule out the possibility of some segregation to native defects, nanometre-scale particles, voids or other interfaces that have been proposed to affect thermoelectric properties, including electronic effects such as electron filtering⁸. However, the success of our model in predicting transport properties using only bulk properties (electronic structure, and electron and phonon scattering) shows that nanoscale effects are not necessary to achieve exceptionally high zT in PbTe alloys (Fig. 2, Supplementary Information). Combining this

b, Lattice thermal conductivity can be well described by an alloy scattering model. Multiple data points in **b** are for different samples with the same Se content but with Na doping varying in the range 1–2 atom %.

exceptional bulk performance with independent nanoscale effects, such as the scattering of long mean free path phonons by nanostructural interfaces^{2–4}, would lead to further enhancements of zT , probably giving values above 2, and also improve the average zT for use in efficient waste heat recovery.

The observed reduction in κ_{L} as Se content increases (Fig. 4b) is expected from alloys, as phonons are scattered due to the mass difference and local strain caused by the impurity atoms, and can be well characterized by the Debye–Callaway model^{4,29} (see Supplementary Information for details). Using the electronic model and the lattice thermal conductivity (κ_{L}), zT can be calculated at any doping level, alloy composition and temperature.

The temperature dependence of the Seebeck coefficient, the resistivity ($\rho = 1/\sigma$) and the thermal conductivity are shown in Fig. 3. The measured temperature dependent transport properties agree very well with the electronic model, confirming that the exceptional thermoelectric properties arise when the valley degeneracy is large, particularly when the L and Σ bands converge (within $\sim 2k_{\text{B}}T$). Because of the high density-of-states mass (m^*) resulting from the convergence of many (at least 12) valleys, heavy doping is required to realize the full potential of high degeneracy to produce high zT . The zT measured on 2% Na-doped $\text{PbTe}_{0.98}\text{Se}_{0.02}$ (1.8 ± 0.1 at 850 K, determined on multiple samples on multiple instruments, Supplementary Fig. 2) shows good agreement with the calculated zT , as seen in Fig. 2.

In summary, high valley degeneracy produced by carrier pocket engineering in a bulk material is an effective strategy to enhance thermoelectric performance through the convergence of conducting electronic bands, provided that the doping is properly tuned. Heavily doped p-PbTe_{1-x}Se_x demonstrates how high valley degeneracy enables high zT , especially when combined with other mechanisms (such as alloy scattering) that reduce κ_{L} . A high zT value of ~ 1.8 at high temperatures make these simple and stable materials superior to those currently in use for thermoelectric energy generation applications.

METHODS SUMMARY

Polycrystalline $\text{Pb}_{0.98}\text{Na}_{0.02}\text{Te}_{1-x}\text{Se}_x$ samples were prepared by melting the mixture of pure elements at 1,273 K, quenching, annealing at ~ 900 K for 3 days, grinding and hot-pressing (98% or higher relative density). X-ray diffraction and scanning electron microscope analyses confirm that the materials for this study were single phase solid solutions. The Seebeck coefficient was obtained from the slope of the thermovoltage versus temperature gradient, confirmed on four different high temperature systems. Scanning Seebeck coefficient measurements (at 300 K) on a sample with a zT of ~ 1.8 at 800 K showed a Seebeck coefficient variation of only $5 \mu\text{V K}^{-1}$ (full width for 90% of the data taken in an area of $6.5 \times 7 \text{ mm}^2$). Four-probe resistivity was measured using the Van der Pauw technique on disks, and using the linear method on bar shaped samples. Thermal diffusivity was measured using the laser flash method. Heat capacity (C_p) is estimated from the relation C_p/k_{B} per atom = $3.07 + (4.7 \times 10^{-4} \times (T - 300))$, where T is in K, based on experimental

literature values³⁰. The combined uncertainty for the experimental determination of zT is $\sim 20\%$; the standard deviation of the measured zT at $T = 800 \pm 50$ K is 4% for four different techniques and 3% for four different samples, all of the same $x = 0.15$ composition. The Hall coefficient at room temperature and higher was measured using the Van der Pauw technique under a reversible magnetic field of ~ 2 T. The low temperature (2.5–300 K) Hall coefficient was measured using a Quantum Design PPMS.

Received 13 December 2010; accepted 14 March 2011.

- Bell, L. E. Cooling, heating, generating power, and recovering waste heat with thermoelectric systems. *Science* **321**, 1457–1461 (2008).
- Kanatzidis, M. G. Nanostructured thermoelectrics: the new paradigm? *Chem. Mater.* **22**, 648–659 (2010).
- Poudeu, P. F. P. *et al.* High thermoelectric figure of merit and nanostructuring in bulk p-type $\text{Na}_{1-x}\text{Pb}_m\text{Sb}_{1-m}\text{Te}_{m+2}$. *Angew. Chem. Int. Edn* **45**, 3835–3839 (2006).
- Pei, Y., Lensch-Falk, J., Toberer, E. S., Medlin, D. L. & Snyder, G. J. High thermoelectric performance in PbTe due to large nanoscale Ag_2Te precipitates and La doping. *Adv. Funct. Mater.* **21**, 241–249 (2011).
- Snyder, G. J. & Toberer, E. S. Complex thermoelectric materials. *Nature Mater.* **7**, 105–114 (2008).
- Mahan, G. D. in *Solid State Physics* (eds Ehrenreich, H. & Spaepen, F.) Vol. 51, 81–157 (Academic, 1998).
- Goldsmid, H. J. *Thermoelectric Refrigeration* (Plenum, 1964).
- Dresselhaus, M. S. *et al.* New directions for low-dimensional thermoelectric materials. *Adv. Mater.* **19**, 1043–1053 (2007).
- Koga, T., Sun, X., Cronin, S. & Dresselhaus, M. Carrier pocket engineering to design superior thermoelectric materials using GaAs/AlAs superlattices. *Appl. Phys. Lett.* **73**, 2950–2952 (1998).
- Rabina, O., Lin, Y. & Dresselhaus, M. Anomalous high thermoelectric figure of merit in $\text{Bi}_{1-x}\text{Sb}_x$ nanowires by carrier pocket alignment. *Appl. Phys. Lett.* **79**, 81–83 (2001).
- Heremans, J. *et al.* Enhancement of thermoelectric efficiency in PbTe by distortion of the electronic density of states. *Science* **321**, 554–557 (2008).
- Slack, G. A. in *CRC Handbook of Thermoelectrics* Ch. 34 (ed. Rowe, D. M.) 406–440 (CRC, 1995).
- DiSalvo, F. J. Thermoelectric cooling and power generation. *Science* **285**, 703–706 (1999).
- Ravich, Y. I., Efimova, B. A. & Smirnov, I. A. *Semiconducting Lead Chalcogenides* (Plenum, 1970).
- Nimtz, G. & Schlögl, B. Narrow-gap lead salts. *Springer Tracts Mod. Phys.* **98**, 27–40 (1983).
- Ravich, Y. I. in *Lead Chalcogenides: Physics and Applications* (ed. Khokhlov, D.) 1–34 (Taylor and Francis, 2003).
- Ashcroft, N. W. & Mermin, N. D. *Solid State Physics* 570 (Brooks Cole, 1976).
- Bilc, D. *et al.* Resonant states in the electronic structure of the high performance thermoelectrics $\text{AgPb}_m\text{SbTe}_{2+m}$: the role of Ag-Sb microstructures. *Phys. Rev. Lett.* **93**, 146403 (2004).
- Tauber, R. N., Machonis, A. A. & Cadoff, I. B. Thermal and optical energy gaps in PbTe. *J. Appl. Phys.* **37**, 4855–4860 (1966).
- Strauss, A. J. Metallurgical and electronic properties of $\text{Pb}_{1-x}\text{Sn}_x\text{Te}$, $\text{Pb}_{1-x}\text{Sn}_x\text{Se}$, and other IV–VI alloys. *Trans. Metall. Soc. AIME* **242**, 354–365 (1968).
- Fritts, R. W. in *Thermoelectric Materials and Devices* (eds Cadoff, I. B. & Miller, E.) 143–162 (Reinhold, 1960).
- Androulakis, J. *et al.* Thermoelectric enhancement in PbTe with K or Na codoping from tuning the interaction of the light- and heavy-hole valence bands. *Phys. Rev. B* **82**, 115209 (2010).
- Pei, Y., LaLonde, A., Iwanaga, S. & Snyder, G. J. High thermoelectric figure of merit in heavy-hole dominated PbTe. *Energy Environ. Sci.* doi:10.1039/c0ee00456a (2011).
- Kudman, I. Thermoelectric properties of p-type PbTe–PbSe alloys. *J. Mater. Sci.* **7**, 1027–1029 (1972).
- Chernik, I. A., Kaidanov, V. I., Vinogradova, M. I. & Kolomoets, N. V. Investigation of the valence band of lead telluride using transport phenomena. *Sov. Phys. Semicond.* **2**, 645–651 (1968).
- Lyden, H. A. Temperature dependence of the effective masses in PbTe. *Phys. Rev. A* **135**, A514–A521 (1964).
- Ravich, Y. I., Efimova, B. A. & Tamarche, V. I. Scattering of current carriers and transport phenomena in lead chalcogenides. 1. Theory. *Phys. Status Solidi B* **43**, 11–33 (1971).
- Crocker, A. J. & Rogers, L. M. Valence band structure of PbTe. *J. Phys. Coll.* **29**, 129–132 (1968).
- Alekseeva, G. T., Efimova, B., Ostrovsk, L. M., Serebrya, O. S. & Tsypin, M. Thermal conductivity of solid solutions based on lead telluride. *Sov. Phys. Semicond.* **4**, 1122–1125 (1971).
- Blachnik, R. & Igel, R. Thermodynamic properties of IV–VI-compounds: leadchalcogenides. *Z. Naturforsch. B* **29**, 625–629 (1974).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by NASA-JPL and the DARPA Nano Materials programme; the work at SIC-CAS was supported by CAS. We thank J.-P. Fleurial, S. Bux, D. Zoltan and F. Harris for measurements of transport properties at NASA's Jet Propulsion Laboratory and at ZT Plus Inc.

Author Contributions Y.P. synthesized the samples, measured the high temperature properties and developed the three-band model; X.S. and L.C. measured the low temperature Hall coefficient and confirmed the high temperature transport properties on some of the samples. A.L. performed the hot pressing; Y.P., X.S., A.L., H.W., L.C. and G.J.S. analysed the experimental data; and Y.P. and G.J.S. wrote and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.J.S. (jsnyder@caltech.edu).

The Soret effect and isotopic fractionation in high-temperature silicate melts

Gerardo Dominguez¹, Gautam Wilkins^{2†} & Mark H. Thiemens¹

Diffusion in condensed phases is a ubiquitous but poorly understood phenomenon. For example, chemical diffusion, which is the transport of matter associated with chemical concentration gradients (Fick's law), is treated as a separate process from thermal transport (the Soret effect), which is mass transport induced by temperature gradients. In the past few years, large variations in the proportions of isotopes of Mg, Ca, Fe, Si and O found in silicate melts subject to thermal gradients have been found^{1–3}, but no physical mechanism has been proposed. Here we present a model of diffusion in natural condensed systems that explains both the chemical and isotopic fractionation of Mg, Ca and Fe in high-temperature geochemical melts. Despite the high temperatures associated with these melts ($T > 1,000^\circ\text{C}$), we find that consideration of the quantum-mechanical zero-point energy of diffusing species is essential for understanding diffusion at the isotopic level. Our model explains thermal and chemical mass transport as manifestations of the same underlying diffusion mechanism. This work promises to provide insights into mass-transport phenomena (diffusion and evaporation) and associated isotopic fractionations in a wide range of natural condensed systems, including the atmospheric water cycle^{1,2}, geological and geochemical systems^{3–6} and the early Solar System⁴. This work might also be relevant to studies of mass transport in biological^{7,8} and nanotechnological condensed systems⁹.

Diffusion in condensed phases, especially in natural systems such as solutions, geochemical melts and solid–solid interfaces, is an important process that remains poorly understood at a quantitative level. Transport induced by thermal gradients is believed to be important in regulating biological processes including DNA replication and possibly the origins of life⁷, as well as in industrial applications^{9,10} and geochemical systems. Therefore, there are good reasons for seeking a quantitative understanding of the underlying physical processes.

In 1879, Charles Soret observed that placing salt solutions in a column with a temperature gradient enhances the chemical concentration of salts at the cold end, something that remains unexplained¹¹. A similar phenomenon is seen in high-temperature ($T > 1,400^\circ\text{C}$) silicate melts, in which elements such as Mg, Ca and Fe become concentrated at the cold end of thermal gradients^{12,13}. Furthermore, large enhancements in the relative concentrations of heavy isotopes have been observed at the cold end of such gradients^{5,6,14–16}. In principle, observations of elemental and isotopic gradients in geochemical melts could provide further insight into the Soret effect, but the lack of a theoretical model prevents this^{10,16}. Empirical descriptions treat the mass flux associated with a thermal gradient as

$$j_T = -D_T \nabla T \quad (1)$$

where D_T is the thermal diffusion coefficient of the diffusing species in the solution and C is the concentration. It is considered a separate process from the mass flux associated with a concentration gradient, which is

$$j_C = -D \nabla C \quad (2)$$

where D is the diffusion constant of the diffusing species. In a steady state the total net flux is zero ($j_T = -j_C$), yielding

$$\nabla C = -S_T \nabla T \quad (3)$$

where S_T is the Soret coefficient ($= D_T/D$), which determines the concentration gradient of diffusing species in the system at steady state.

Natural silicate melts and silicate glasses are composed of a network of SiO_4 polyhedra, which form the structural backbone of the condensed phase¹⁷. Secondary elements such as Mg, Ca and Fe are believed to diffuse through interstitial sites, in which electrostatic interactions with surrounding atoms in the network minimize their potential energy (U). Silicates at high temperatures ($T > 1,000\text{ K}$) seem to retain, to a good approximation, their room-temperature (300 K) structural characteristics¹⁷. However, the rate at which species in silicate melts and crystals diffuse over restricted temperature regions increases exponentially as a function of temperature, displaying Arrhenius-like behaviour of the form

$$D(T) = D_0 \exp\left(\frac{-E_A}{k_B T}\right) \quad (4)$$

where E_A is the activation energy for diffusion, k_B is the Boltzmann constant and D_0 is a temperature-independent pre-factor¹⁸. This temperature dependence strongly implies that diffusion in high-temperature silicate melts, like many chemical reactions, is thermally activated.

A theoretical understanding of the temperature dependence of chemical reactions was first provided by transition-state theory. In essence, this theory states that the rate of chemical reactions ($A + B \rightarrow AB$) is determined by the rate at which an activated and unstable transition state (AB^*) is populated and proceeds to the product state ($A + B \rightarrow AB^* \rightarrow AB$; ref. 19). This rate is given by

$$k^{\text{TST}} = \left(\frac{k_B T}{h}\right) e^{-E_A/k_B T} \quad (5)$$

where h is Planck's constant. In this case, the activation energy, E_A , is the free energy difference between the reactant state (RS) and the transition state (TS), which is in turn given by

$$E_A = \Delta E_e - k_B T \ln\left(\frac{Z(\text{TS})}{Z(\text{RS})}\right) \quad (6)$$

where ΔE_e is the electronic energy difference and $Z(\text{RS})$ and $Z(\text{TS})$ represent the partition functions of the RS and TS respectively²⁰. Although transition-state theory has had some success in quantitatively describing diffusion in crystalline materials^{21,22}, its application to diffusion phenomena in natural condensed systems has been limited. Here we show how transition-state theory can be used to understand the chemical and isotopic fractionation of elements in high-temperature silicate melts subject to a thermal gradient.

Consider elemental diffusion in a silicate melt. At the atomic level, these species, it is widely believed, move by occupying interstitial sites in the silicate network^{23–25}, which represent local minima in the electronic energy landscape established by the network of Si–O bonds and

¹University of California, San Diego, Department of Chemistry and Biochemistry, California, 92093, USA. ²University of California, Berkeley, Department of Mathematics, 94720, California, USA. [†]Present address: University of California, San Diego, Department of Chemistry and Biochemistry, California, 92093, USA.

the Coulomb interactions of the diffusing species with this network. The energy to thermally 'hop' out of these sites is provided by random thermal fluctuations (at a temperature T), so this hopping process can be treated thermodynamically. In this view, the diffusion of elements in silicate melts is limited by the ability of diffusing species to overcome local energetic barriers (with electronic energy ΔE_e). The initial location of a diffusing species can be represented as the reactant state, and the local maxima (saddle points) that separate local minima in the network can be thought of as the activated, or transition, states of the diffusing species (See Fig. 1a). The rate of diffusion predicted by transition-state theory (D^{TST}) can be written as

$$D^{\text{TST}} = f_G a_0^2 k^{\text{TST}} \quad (7)$$

where f_G is a geometric structure factor that describes the number of neighbouring sites that a diffusing species can occupy, and a_0 is the distance between jumps (the jump length), which would be on the order of the lattice parameter of the system²¹.

In the most rigorous sense, the partition functions of the entire N-body silicate system when a diffusing species is in the reactant and transition states is needed to calculate k^{TST} (ref. 22). Here we follow Wert and Zener's simplified approach²¹, and write the rate of diffusion as

$$D^{\text{TST}}(m, T) = f_G a_0^2 \left(\frac{k_B T}{h} \right) 2 \sinh \left(\frac{h\nu(m)^{\text{RS},1}}{2k_B T} \right) e^{-\Delta E_e/k_B T} \quad (8)$$

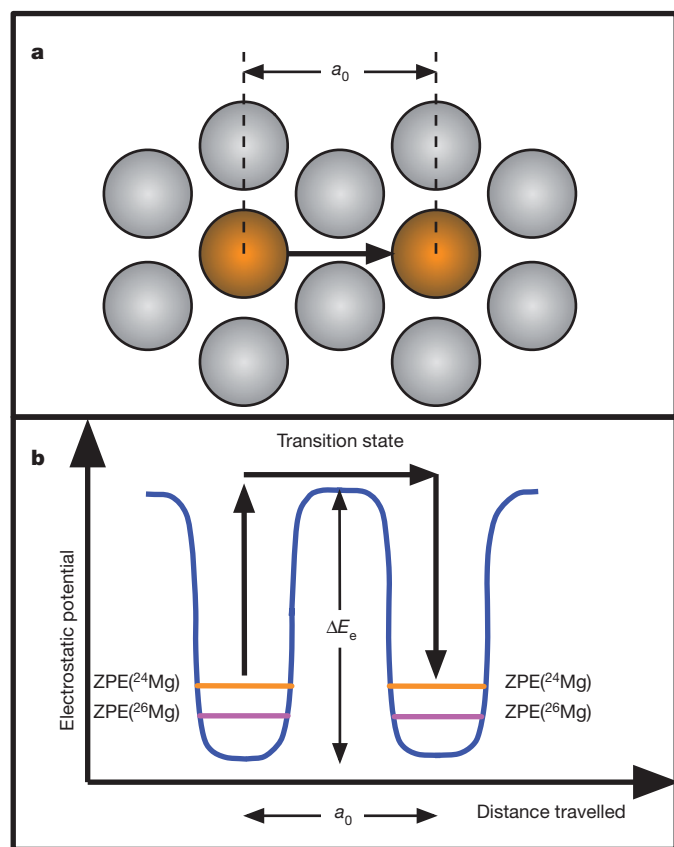


Figure 1 | Diffusion of elements as a random hopping process in a silicate melt. **a**, The diffusing element is shown in orange; the silicate melt is grey. Formation of the transition state, which has a higher potential energy than the element in the interstitial site, is the rate-limiting step in diffusion. The jump length (a_0) is a characteristic length separating interstitial sites in the silicate melt. **b**, The corresponding energetic landscape seen by diffusing species. The zero-point energies of ^{24}Mg and ^{26}Mg are indicated as $\text{ZPE}(^{24}\text{Mg})$ and $\text{ZPE}(^{26}\text{Mg})$ to illustrate the origin of the isotopic mass-dependence of diffusion in condensed phases. ΔE_e , difference in electronic energy.

where $\nu(m)^{\text{RS},1}$ ($= \sqrt{\frac{\kappa}{m}}$) is the vibrational frequency of the diffusing species along the direction of diffusion (x), m is the mass of the diffusing species, and the effective spring constant (κ) is given by the second derivative of the electronic potential energy (U) with respect to x , evaluated at the coordinate (x_{min}) where $U(x)$ is minimized, so that

$$\kappa = \left. \frac{\partial^2 U}{\partial x^2} \right|_{(x=x_{\text{min}})} \quad (9)$$

Equation (8) is intrinsically quantum mechanical because it results from treating the energy levels and the partition function of the diffusing species (Z_v) in the reactant state as those of a quantum harmonic oscillator, which is to say that²⁶:

$$Z_v = \frac{1}{2 \sinh \left(\frac{h\nu}{2k_B T} \right)}$$

Equation (8) also explains why there are differences in the rates of diffusion of isotopes of any given element in a condensed system. It may be tempting, given our eventual goal of describing diffusion in high-temperature geochemical systems, to take the classical limit of these partition functions, but for now we will use the exact expressions and defer any approximations until we have a quantitative estimate of $\frac{h\nu(m)^{\text{RS},1}}{2k_B T}$. Our analysis will show that the quantum-mechanical treatment is needed to accurately describe diffusion in high-temperature condensed phases at the isotopic level. Furthermore, it will show that the dominant effect that determines the relative rate of diffusion for isotopes of an element is the differences in their masses.

Our approach in adapting transition-state theory to elemental diffusion in high-temperature geochemical systems was to use the empirical data on the chemical and isotopic fractionations of Mg from the experiments described in refs 5 and 6 to deduce reasonable values for ΔE_e and E_A in high-temperature silicate melts. To do this, we performed a set of numerical simulations of diffusion, as described in the Methods.

Using this approach, we inferred the values for ΔE_e and $\nu(m)^{\text{RS},1}$ of Mg isotopes that provided the best fit for the isotopic fractionation of Mg in high-temperature silicate melts. We used these results to predict the isotopic fractionation of Ca and Fe in high-temperature silicate melts.

We found that the electronic energy barrier that best reproduced the observed isotopic fractionations of ^{26}Mg was $\Delta E_e = 2.2\text{--}2.5$ eV ($f_G = 0.5\text{--}3$). These electronic energies, in turn, suggest that $E_A = 1.76\text{--}2.0$ eV, in excellent agreement with the activation energies found for elemental Mg diffusion (1.8–2.1 eV) in a basalt melt¹⁸ (see Supplementary Table 1, Supplementary Information). We found that the isotopic fractionation gradients are most sensitive to differences in the activation energies of isotopes of an element.

The vibrational zero-point energies (ZPEs) of roughly 0.4 eV are, at first glance, surprisingly high, because they suggest that the ZPE of diffusing Mg ions is about 20% of ΔE_e . To evaluate whether these ZPEs are physically realistic, we considered some estimates of the spring constant (and corresponding vibrational ZPEs) in silicate melts. Using equation (9), we get

$$\kappa \approx \frac{2\Delta E_e}{L_c^2} \quad (10)$$

for the spring constant, where L_c is the characteristic length scale that describes how steeply the electronic potential energy U changes. The corresponding ZPE is

$$\text{ZPE}(m) = h \sqrt{\frac{\Delta E_e}{2mL_c^2}} \quad (11)$$

where $\text{ZPE}(m)$ is the ZPE for a diffusing species of mass m . A lower-limit estimate of ZPE for Mg is given by setting L_c equal to Si–O bond length²⁷; this length scale underestimates the ZPE that we infer from

the isotopic fractionations by a factor of roughly 4 (ZPE of about 0.1 eV versus 0.4 eV). To explain the large isotopic fractionations observed experimentally for Mg isotopes, we find that the relevant L_c is roughly 0.2 Å (where the numerical value for the spring constant is $\kappa^* \sim 1,200 \text{ J m}^{-2}$), which is not unreasonable, considering that this scale corresponds to about 15% of the ionic radius of oxygen (which is roughly 1.35 Å) in a silicate melt¹⁸. This suggests that quantum confinement of Mg ions (and, by extension, others) in interstitial sites determines their rate of diffusion and, ultimately, their rate of isotopic fractionation during diffusive processes. This length scale is also in qualitative agreement with potential-energy curves of interactions in glass melts from molecular dynamics simulations²⁷.

Because the ZPE of Mg (and, by extension, other elements) in silicate melts is large compared with the thermal energy, the diffusion constants of these elements (equation (5)) and their isotopes can be written to a good approximation as

$$D^{\text{TST}}(m, T) \approx f_G a_0^2 \left(\frac{k_B T}{h} \right) e^{-(\Delta E_c - \text{ZPE}(m))/k_B T} \quad (12)$$

We now revisit Soret diffusion in light of our findings. Adopting a generalized Fokker–Planck diffusivity law in one dimension^{9,28,29}, such that

$$j^{\text{total}}(x, t) = \frac{d}{dx} (D(x) C(x, t)) \quad (13)$$

and assuming that for high-temperature silicate systems there is no significant change in the electronic energy barrier (which would be evidenced by a sharp discontinuity in the chemical concentration gradients between the hot and cold ends, $\frac{d(\Delta E_c)}{dx} = 0$), we can express the total mass flux in a one-dimensional thermal gradient as:

$$j^{\text{total}}(x, t) = C(x, t) \frac{dD}{dx} + D(x) \frac{dC(x)}{dx} = C(x, t) \frac{dD(x)}{dT} \frac{dT}{dx} + D(x) \frac{dC(x)}{dx} \quad (14)$$

Comparing this expression with equations (1) and (2), we can now see that the first and second terms on the right are the mass fluxes associated with thermal and concentration gradients, respectively. This clearly shows that thermal diffusion is the result of the temperature (and therefore spatial) dependence of the diffusion ‘constant’ $D(T)$ in Soret experiments. Substituting expression (12) for $D(T)$, the total flux is now given by

$$j^{\text{total}}(x, t) = C(x, t) \left[1 + \frac{\Delta E_c - \text{ZPE}(m)}{k_B T} \right] \left(\frac{D(x)}{T} \right) \frac{dT}{dx} + D(x) \frac{dC(x)}{dx} \quad (15)$$

and the Soret coefficient ($= D_T/D$) for diffusion in the high-temperature melt is given by

$$S_T(m, T) = - \left[1 + \frac{\Delta E_c - \text{ZPE}(m)}{k_B T} \right] \left(\frac{1}{T} \right) \quad (16)$$

The above expression makes a direct connection between the microphysics of diffusion in a condensed phase and the macroscopically observable elemental and isotopic fractionations of these elements in a thermal gradient. Most relevant for estimating the isotopic fractionation¹⁶ is the difference in the Soret coefficient of two isotopes of an element. This is given by

$$S_T(m_1, T) - S_T(m_2, T) = \Delta S_T(m_1, m_2) = - \left[\frac{h\sqrt{\kappa}}{2} \left(\frac{1}{\sqrt{m_1}} - \frac{1}{\sqrt{m_2}} \right) \right] \left(\frac{1}{k_B T^2} \right) \quad (17)$$

This expression is, to first order, independent of the electronic energy, and is a function of the relative difference between the masses of the isotopes. It correctly predicts larger isotopic variations for lighter isotopes and elements. As shown by ref. 16, differences in the Soret

coefficient lead to the establishment of isotopic gradients. Our model predicts that the isotopic difference between the hot (T_H) and cold (T_L) end of a thermal gradient is given by

$$\delta^X M(T_L) - \delta^X M(T_H) = \left[\frac{h\sqrt{\kappa}}{k_b} \left(\frac{1}{\sqrt{m_1}} - \frac{1}{\sqrt{m_2}} \right) \right] \left(\frac{1}{T_L} - \frac{1}{T_H} \right) \quad (18)$$

The isotopic fractionations for Mg predicted using the above approximate expression (with κ^*) agreed with our numerical simulations to within 5%.

We provide further confirmation of our model by taking the spring constant κ^* and using equation (18) to predict the isotopic fractionation for ²⁵Mg, ⁵⁶Fe, ⁵⁷Fe and ⁴⁴Ca under the same conditions as ²⁶Mg. These comparisons (Fig. 2a–e) show that our model reproduces the empirical observations of isotopic fractionation of Mg, Ca and Fe (refs 5, 6 and 16) very well. However, we find that our model overpredicts the fractionation of Si and O if we assume that they diffuse as ionic species (Fig. 2f). If we assume that Si diffuses as SiO₄ tetrahedra, the agreement is excellent. In contrast, if we assume that O diffuses as SiO₄ tetrahedra, our model underpredicts the observations, giving them the same slope as for Si. This suggests that oxygen diffuses as both SiO₄ and a less massive species (atomic O).

Our model of diffusion accurately predicts the direction (sign) of the Soret effect for Mg, Ca and Fe, and the general magnitude of the effect. The concentration gradients themselves have a wide range of magnitudes, and in Supplementary Table 2 we compare the observed values for S_T with those predicted by our model. The agreement is fair to excellent, especially in light of the simplicity of the proposed model. The differences, where they exist, may not be too surprising, because our model assumes the absence of an electronic energy gradient between the cold and hot ends of the thermal gradient. The enrichment of species such as SiO₂ at the hot end of Soret experiments is likely to affect the electronic energy landscape seen by diffusing species. To evaluate the importance of this issue, we performed a sensitivity analysis of our general expression for the Soret coefficient S_T (See Supplementary Table 3 and discussion in Supplementary Information). We found that small differences (1–10%) in the electronic or activation energy can induce rather large changes in the chemical concentration

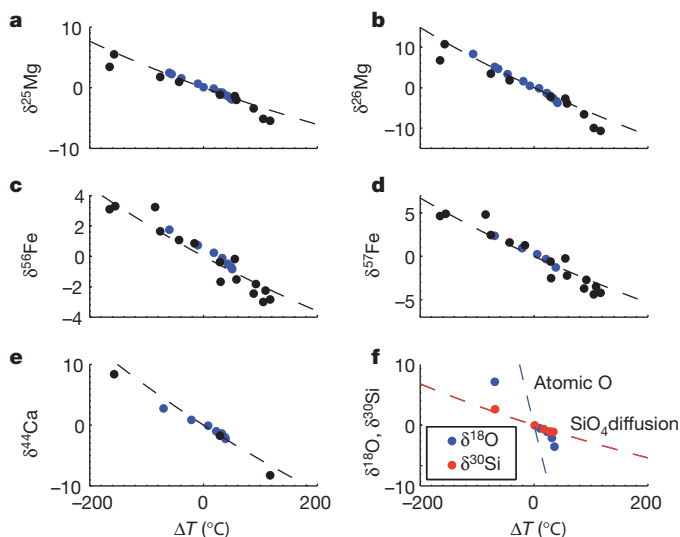


Figure 2 | Model predictions and observations of isotopic fractionation of Mg, Fe, Ca, Si and O in silicate melts (a–f). Numerical simulations of Mg isotope diffusion in a thermal gradient ($T_0 = 1,470^\circ\text{C}$, $\Delta T = 100^\circ\text{C}$) were used to find κ^* ($\sim 1200 \text{ J m}^{-2}$). Using equation (18) and κ^* , we calculated the predictions of our model (dashed line) and compared these to the empirical observations of refs 5 and 6 (blue dot and red dot) and ref. 16 (black dot) for (a) ²⁵Mg/²⁴Mg, (b) ²⁶Mg/²⁴Mg, (c) ⁵⁶Fe/⁵⁴Fe (d) ⁵⁷Fe/⁵⁴Fe, (e) ⁴⁴Ca/⁴⁰Ca, (f) ³⁰Si/²⁸Si and ¹⁸O/¹⁶O.

gradients of elements, consistent with the observation that concentration gradients in Soret experiments are very sensitive to experimental conditions. The near-constant value of the isotopic gradients in the high-temperature Soret diffusion experiments we have analysed here suggest that the ZPE differences of isotopes are less sensitive to changes in chemical composition.

Although our model does not predict the direction of diffusion for cations such as Si^{4+} , Na^+ and K^+ , we have shown through sensitivity analysis (see Supplementary Information and Supplementary Table 3) how the direction of flow in thermal gradients is extremely sensitive to gradients in the electronic energy (equivalent to chemical potential gradients). Future work should clarify how chemical composition and electronic energy potential energy relate to one another.

We have developed an atomistic model for diffusion in natural condensed systems, and have shown that it can be used to understand the chemical and isotopic fractionation of elements in high-temperature silicate melts. We find that relative differences in the masses of isotopes of an element determine their isotopic fractionation. Because of its very general nature, this model and extensions of it will find applications and lead to new insights in understanding elemental and isotopic diffusion in other natural systems, such as the global hydrological cycle^{1,2}, the isotopic fractionations associated with the evaporation of early Solar System materials⁴, and transport in industrial processes⁹. Furthermore, the model that we have presented, together with measurements of the spatial distribution of isotopic ratios, may be used in geological systems to reconstruct their cooling histories^{3–6}.

METHODS SUMMARY

We performed numerical simulations of diffusion in the silicate melts subjected to a thermal gradient using a one-dimensional finite grid. Each of these sheets was assumed to be in local thermodynamic equilibrium, and the temperatures of each sheet were chosen to mimic the thermal gradients of experiments. The concentration of each isotopic species of mass m_i was initially set to the species's natural abundance, and the numerical flux ($dN(m_i)/dt$) into and out of each sheet was computed (See Methods). This approach results in the establishment of a set of coupled differential equations whose time-evolution was solved using an ordinary differential equation solver in Matlab. The numerical concentrations in each of the sheets were then used to calculate the ratios of minor to major isotopes, and were expressed using standard δ notation (see Methods).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 October 2010; accepted 9 February 2011.

Published online 20 April 2011.

1. Luz, B., Barkan, E., Yam, R. & Shemesh, A. Fractionation of oxygen and hydrogen isotopes in evaporating water. *Geochim. Cosmochim. Acta* **73**, 6697–6703 (2009).
2. Uemura, R., Barkan, E., Abe, O. & Luz, B. Triple isotope composition of oxygen in atmospheric water vapor. *Geophys. Res. Lett.* **37**, L04402 (2010).
3. Richter, F. M., Dauphas, N. & Teng, F.-Z. Non-traditional fractionation of non-traditional isotopes: evaporation, chemical diffusion and Soret diffusion. *Chem. Geol.* **258**, 92–103 (2009).
4. Richter, F. M., Janney, P. E., Mendybaev, R. A., Davis, A. M. & Wadhwa, M. Elemental and isotopic fractionation of Type B CAI-like liquids by evaporation. *Geochim. Cosmochim. Acta* **71**, 5544–5564 (2007).
5. Richter, F. M. *et al.* Isotopic fractionation of the major elements of molten basalt by chemical and thermal diffusion. *Geochim. Cosmochim. Acta* **73**, 4250–4263 (2009).

6. Richter, F. M., Watson, E. B., Mendybaev, R. A., Teng, F.-Z. & Janney, P. E. Magnesium isotope fractionation in silicate melts by chemical and thermal diffusion. *Geochim. Cosmochim. Acta* **72**, 206–220 (2008).
7. Braun, D. & Libchaber, A. Thermal force approach to molecular evolution. *Phys. Biol.* **1**, 1–8 (2004).
8. Korkotian, E. & Segal, M. Spatially confined diffusion of calcium in dendrites of hippocampal neurons revealed by flash photolysis of caged calcium. *Cell Calcium* **40**, 441–449 (2006).
9. Duhr, S. & Braun, D. Thermophoretic depletion follows Boltzmann distribution. *Phys. Rev. Lett.* **96**, 168301 (2006).
10. Duhr, S. & Braun, D. Why molecules move along a temperature gradient. *Proc. Natl Acad. Sci. USA* **103**, 19678–19682 (2006).
11. Soret, C. Sur l'état d'équilibre que prend au point de vue de sa concentration une dissolution saline primitivement homogène dont deux parties sont portées à des températures différentes. *Arch. Sci. Phys. Nat.* **2**, 48–61 (1879).
12. Walker, D. & DeLong, S. E. Soret separation of mid-ocean ridge basalt magma. *Contrib. Mineral. Petrol.* **79**, 231–240 (1982).
13. Leshner, C. E. & Walker, D. Solution properties of silicate liquids from thermal diffusion experiments. *Geochim. Cosmochim. Acta* **50**, 1397–1411 (1986).
14. Huang, F. Chemical and isotopic fractionation of wet andesite in a temperature gradient: experiments and models suggesting a new mechanism of magma differentiation. *Geochim. Cosmochim. Acta* **73**, 729–749 (2009).
15. Huang, F. Magnesium isotopic composition of igneous rock standards measured by MC-ICP-MS. *Chem. Geol.* **268**, 15–23 (2009).
16. Huang, F. *et al.* Isotope fractionation in silicate melts by thermal diffusion. *Nature* **464**, 396–400 (2010).
17. Vashishta, P., Kalia, R. K. & Rino, J. P. Interaction potential for SiO_2 : A molecular-dynamics study of structural correlations. *Phys. Rev. B* **41**, 12197–12209 (1990).
18. LaTourrette, T., Wasserburg, G. J. & Fahey, A. J. Self diffusion of Mg, Ca, Ba, Nd, Yb, Ti, Zr, and U in haplobasaltic melt. *Geochim. Cosmochim. Acta* **60**, 1329–1340 (1996).
19. Truhlar, D. G., Garrett, B. C. & Klippenstein, S. J. Current status of transition-state theory. *J. Phys. Chem.* **100**, 12771–12800 (1996).
20. Eyring, H. The activated complex in chemical reactions. *J. Chem. Phys.* **3**, 107–115 (1935).
21. Wert, C. & Zener, C. Interstitial atomic diffusion coefficients. *Phys. Rev.* **76**, 1169–1175 (1949).
22. Vineyard, G. H. Frequency factors and isotope effects in solid state rate processes. *J. Phys. Chem. Solids* **3**, 121–127 (1957).
23. Roselieb, K. & Jambon, A. Tracer diffusion of potassium, rubidium, and cesium in a supercooled jadeite melt. *Geochim. Cosmochim. Acta* **61**, 3101–3110 (1997).
24. Chakraborty, S. Diffusion in solid silicates: a tool to track timescales of processes comes of age. *Annu. Rev. Earth Planet. Sci.* **36**, 153–190 (2008).
25. Watson, E. B. & Baxter, E. F. Diffusion in solid-Earth systems. *Earth Planet. Sci. Lett.* **253**, 307–327 (2007).
26. Pathria, R. K. *Statistical Mechanics*, 2nd ed. (Butterworth-Heinemann, 1996).
27. Kubicki, J. D. & Lasaga, A. C. Molecular dynamics simulations of SiO_2 melt and glass: ionic and covalent models. *Am. Mineral.* **73**, 941–955 (1988).
28. Landsberg, P. T. $D \text{ grad } v$ or $\text{grad}(Dv)$? *J. Appl. Phys.* **56**, 1119–1122 (1984).
29. van Milligen, B. P., Bons, P. D., Carreras, B. A. & Sanchez, R. On the applicability of Fick's law to diffusion in inhomogeneous systems. *Eur. J. Phys.* **26**, 913–925 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Frank Richter for introducing us to this problem. We thank C. Lundstrom, E. Schauble, S. Chakraborty, A. Fnu and R. Shaheen for discussions. G.D. acknowledges support from the University of California President's Fellowship and NASA grants NNX07AM66G and NNX08A15G.

Author Contributions G.D. led the theoretical work. G.W. and G.D. developed and performed numerical simulations, and G.D., G.W. and M.H.T. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.D. (gdominguez@ucsd.edu).

METHODS

To determine physically realistic values for the activation energies, we set the jump length to the Si–O distance of silicate melts ($a_0 = 1.6 \text{ nm}$)^{17,27}. The values of f_G (about 1) and ΔE_e (1–4 eV) were held fixed, and the vibrational frequency of the diffusing species ($\nu(m)^{\text{RS},1}$) was determined by equating D^{TST} for ^{24}Mg , using equation (7) with the empirical value for the diffusion constant of elemental Mg at 1480°C ($D = 7 \times 10^{-7} \text{ cm}^2\text{s}^{-1}$) in a basalt melt.

The vibrational frequency of an isotope of mass m_1 trapped in a harmonic potential is given by

$$\nu(m_1) = \sqrt{\frac{\kappa}{m_1}} \quad (19)$$

Chemically, isotopes of an element are expected to interact identically with their environments; that is, they have the same ΔE_e and L_c , and therefore the same κ . This allows us to directly calculate the ratio of the vibrational frequencies (and corresponding ZPEs) of two isotopes as

$$\frac{\nu(m_1)}{\nu(m_2)} = \sqrt{\frac{m_2}{m_1}} \quad (20)$$

The vibrational frequency for the minor isotope, ^{26}Mg , was determined by the scaling relationship

$$\frac{\nu(m_{26})}{\nu(m_{24})} = \sqrt{\frac{m_{24}}{m_{26}}}$$

as is expected under the harmonic-oscillator approximation.

We used a one-dimensional finite grid model that assumes local thermodynamic equilibrium^{9,10,30} to account for the mass fluxes associated with both chemical concentration and thermal gradients at a nanoscopic level. The silicate melt was represented as a finite set of planes with position x_i . For each of these planes, a local temperature $T(x_i)$, which ranged from 1300°C to 1520°C , was set to mimic the thermal gradients of the experiments in refs 5 and 6.

The mass-flux rate per unit area out of a given sheet (j^-) for a species of mass m_j was computed as

$$j^-(x_i, m_j, t) = -C(x_i, m_j, t) \times k^{\text{TST}}(T(x_i), m_j) \quad (21)$$

where $C(x_i, m_j)$ is its concentration at position x_i , and $k^{\text{TST}}(T(x_i), m_j)$ was extracted from equation (8). The mass flux into a sheet (j^+) (except at the edges of the system), was computed as

$$j^+(x_i, m_j, t) = \frac{C(x_{i-1}, m_j, t) \times k^{\text{TST}}(T(x_{i-1}), m_j)}{2} + \frac{C(x_{i+1}, m_j, t) \times k^{\text{TST}}(T(x_{i+1}), m_j)}{2} \quad (22)$$

and the total mass flux was computed as $j^{\text{total}} = j^+ + j^-$. The concentration of isotopic species in the silicate melt as a function of time and position was determined by solving a set of coupled linear differential equations (one for each sheet i) using an ordinary differential equation solver in Matlab (ODE15S). As inputs, we provided this strike for initial numerical concentrations of each isotope for each grid point

$$N(x_i, m_j, t=0) = C(x_i, m_j, 0) \times A \times a_0 \quad (23)$$

where the area of each sheet (A) is set equal to 1 and a_0 is the thickness of the layer (= jump length). The rate of change of this numerical concentration is given by

$$\frac{dN(x_i, m_j, t)}{dt} = [j^+(x_i, m_j, t) + j^-(x_i, m_j, t)] \times A \times a_0 \quad (24)$$

The isotopic composition as a function of position and time was reported as $\delta^{26}\text{Mg}$ (see below). Initially, the concentrations of ^{24}Mg and ^{26}Mg in the silicate melt were assumed to be uniform and at natural abundances. Quality control during the integrations was set by ensuring that the total number of isotopic species was conserved throughout the simulations.

Isotopic ratio quantification. The isotopic compositions of ^{25}Mg and ^{26}Mg are reported with respect to ^{24}Mg as

$$\delta^{25,26}\text{Mg}(x, t) = \left[\frac{\left(\frac{N(x, m_{25,26}, t)}{N(x, m_{24}, t)} \right)}{\left(\frac{N(x, m_{25,26}, t=0)}{N(x, m_{24}, t=0)} \right)} - 1 \right] \times 10^3 \quad (25)$$

For ^{44}Ca isotopes as

$$\delta^{44}\text{Ca}(x, t) = \left[\frac{\left(\frac{N(x, m_{44}, t)}{N(x, m_{40}, t)} \right)}{\left(\frac{N(x, m_{44}, t=0)}{N(x, m_{40}, t=0)} \right)} - 1 \right] \times 10^3 \quad (26)$$

And for ^{56}Fe and ^{57}Fe as

$$\delta^{56,57}\text{Fe}(x, t) = \left[\frac{\left(\frac{N(x, m_{56}, t)}{N(x, m_{54}, t)} \right)}{\left(\frac{N(x, m_{56,57}, t=0)}{N(x, m_{54}, t=0)} \right)} - 1 \right] \times 10^3 \quad (27)$$

where $n(z, m_j, t)$ represents the concentration of isotope of mass m_x at position x and time t . Initially, these isotope ratios are equal to 0 by definition; the subsequent isotopic composition shifts are insensitive to the choice of normalization.

30. Astumian, R. D. Coupled transport at the nanoscale: the unreasonable effectiveness of equilibrium theory. *Proc. Natl Acad. Sci. USA* **104**, 3–4 (2007).

Depth-dependent extension, two-stage breakup and cratonic underplating at rifted margins

Ritske Huismans¹ & Christopher Beaumont²

Uniform lithospheric extension¹ predicts basic properties of non-volcanic rifted margins but fails to explain other important characteristics^{2,3}. Significant discrepancies are observed at ‘type I’ margins (such as the Iberia–Newfoundland conjugates), where large tracts of continental mantle lithosphere are exposed at the sea floor⁴, and ‘type II’ margins (such as some ultrawide central South Atlantic margins), where thin continental crust spans wide regions below which continental lower crust and mantle lithosphere have apparently been removed^{5,6}. Neither corresponds to uniform extension. Instead, either crust or mantle lithosphere has been preferentially removed. Using dynamical models, we demonstrate that these margins are opposite end members: in type I, depth-dependent extension results in crustal-necking breakup before mantle-lithosphere breakup and in type II, the converse is true. These two-layer, two-stage breakup behaviours explain the discrepancies and have implications for the styles of the associated sedimentary basins. Laterally flowing lower-mantle cratonic lithosphere may underplate some type II margins, thereby contributing to their anomalous characteristics.

Passive margins produced by continental rifting and ocean-floor spreading have a wide range of characteristics, many of which remain enigmatic. We focus on two styles of non-volcanic (magma-poor) margins, which we term types I and II. We list their primary characteristics and show that they are end members with respect to the way the lithosphere stretched as they formed. Type I margins (Fig. 1a), such as the Iberia–Newfoundland conjugate margins^{7–9}, usually develop after distributed extension, which finally becomes focused on one location¹⁰.

The defining characteristics of the focused regions of type I (Fig. 1a and 2a–c) include⁹: major basin-forming faults or shears that penetrate deep into the crust (characteristic 1); narrow regions (less than about 100 km wide) across which the crust thins abruptly (2); usually an asymmetric geometry and uplift of rift flanks (3); breakup of crust before that of the mantle lithosphere (4); exhumation and exposure of serpentinized continental mantle lithosphere in the ocean–continent transition (5); limited magmatism during rifting, leading to a magma-poor margin (6); and delayed establishment of an oceanic spreading centre and normal ocean crust production (7).

In contrast, type II margins (Fig. 1b), as exemplified by some margins in the central South Atlantic^{5,6} and the Exmouth plateau¹¹, and in particular those shown in Fig. 3 and Supplementary Fig. 3, have different characteristics. These are: ultrawide regions of thin continental crust (characteristic A); faulted early syn-rift sedimentary basins (B); undeformed late syn-rift sediments (C); the capping of these late syn-rift sediments by evaporites and other sediments, deposited in shallow water conditions in ‘sag’ basins (D); limited syn-rift subsidence owing to replacement of underlying continental mantle lithosphere by hot asthenosphere, as suggested by these sag basins (E); no syn-rift flank uplifts (F); no clear evidence of exposed mantle lithosphere, but some syn-rift magmatism (G); lower-crustal regions with seismic velocities consistent with magmatic underplating (H); and a normal magmatic mid-oceanic-ridge/crust system established soon after crustal breakup (I).

We propose that the characteristics of type I margins (Fig. 1a) are explained by reference to model I (Fig. 2), which demonstrates a style of depth-dependent extension in which the crust and mantle lithosphere are strong and strongly bonded (Supplementary Fig. 6c). It has a two-layer rheology in which the upper lithosphere undergoes frictional–plastic (brittle) deformation while the lower lithosphere deforms by viscous (ductile) power-law flow (Supplementary Methods). Rifting comprises three phases (Supplementary Movie 1). First, subsidence of a symmetric keystone crustal block bounded by conjugate frictional–plastic faulting or shearing of the upper layer, underlain by ductile necking of the lower lithosphere (Fig. 2d); followed by asymmetric extension resembling conceptual simple shear¹² (Fig. 2e); and finally rupture of the crust, followed by continued extension, necking, exhumation and exposure of the lower lithosphere. Small variations in conditions make the asymmetry more or less pronounced.

Comparison of model I (Fig. 2d–f) with type I margins (Figs 1 and 2a–c) shows consistent characteristics, namely the type I primary characteristics (1) to (6) listed above⁹. In particular, major basin-forming conjugate faults and shears that penetrate at least to the lower-crust (1), narrow transitional regions (<100 km wide) of crustal thinning (2) (Fig. 2b and c), asymmetry (3) (Fig. 2b and c) and exhumation of continental mantle lithosphere (5). The model I results also explain one additional puzzling characteristic of the Iberia–Newfoundland conjugates¹³: extreme crustal thinning after only limited extension⁹ is achieved by superimposed conjugate and detachment shearing (Fig. 2d and e).

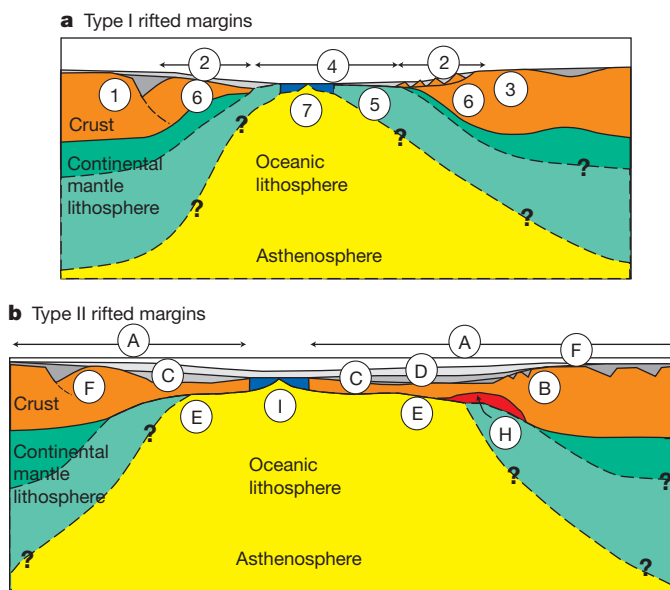


Figure 1 | Characteristic properties of type I and type II margins. Type I (a) and type II (b) rifted continental margins based on observations from the Iberia–Newfoundland conjugate margins and central South Atlantic margins, respectively^{4–10,13–17,31}. See text for their typical characteristics 1–7 and A–I.

¹Department of Earth Science, Bergen University, Bergen, N-5007, Norway. ²Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, B3H 4J1, Canada.

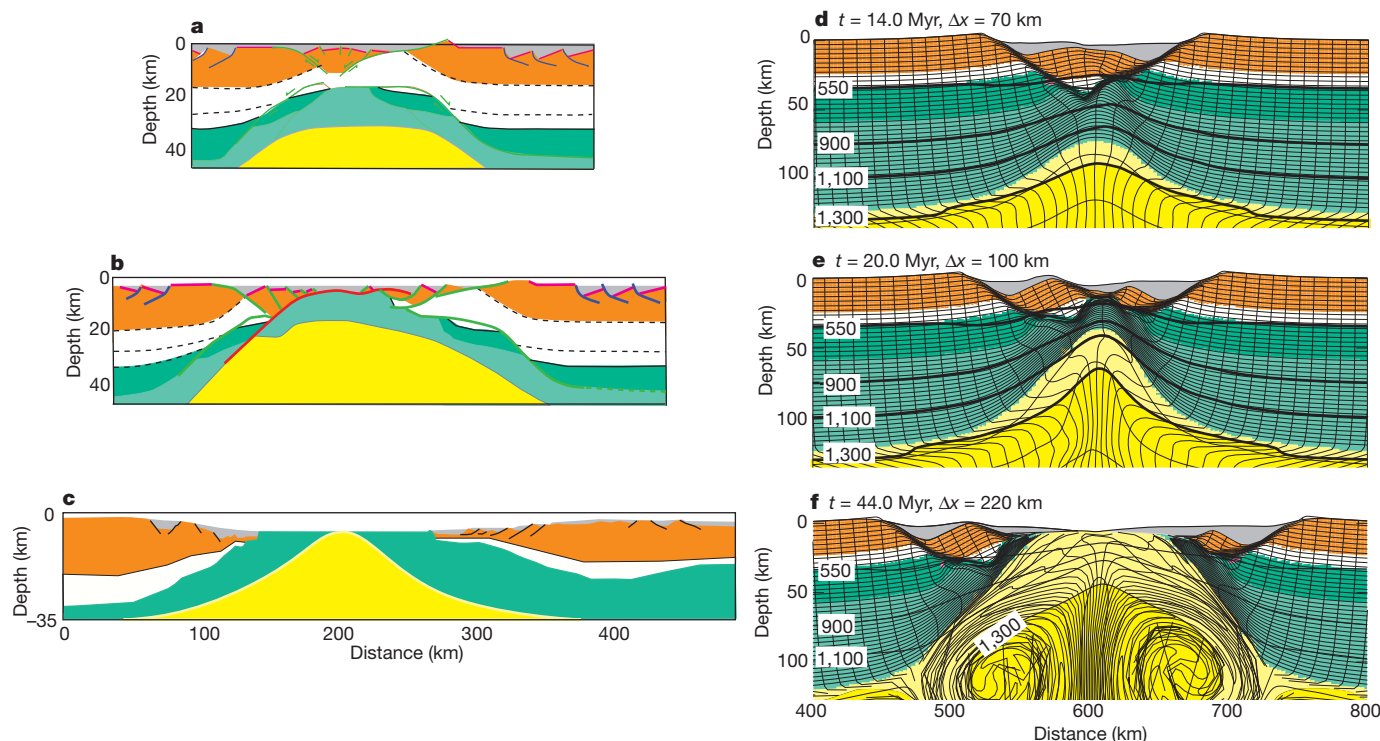


Figure 2 | Type I margins. **a, b**, Conceptual reconstructions^{9,10} of the Iberia-Newfoundland conjugate margins (see Supplementary Information). **c**, Interpreted observations⁷⁻⁹ restored to late Aptian. Compare with model I results. **d-f**, Model I results (shown for a subregion of the model domain; see Supplementary Movie 1). Myr, millions of years; t , time since the onset of

extension; Δx , extension at uniform velocity 0.5 cm yr^{-1} . Contours are isotherms in degrees Celsius. Shown are sediments (grey), upper and mid-crust (orange), lower crust (white), frictional-plastic (dark green) and viscous (green) continental mantle lithosphere, oceanic lithosphere (pale yellow) and asthenosphere (yellow).

The characteristics of type II margins (Fig. 1b), as listed above and in Supplementary Information 2, are consistent with model II-A (where A indicates asthenospheric inflow). This model allows decoupling between the upper and lower lithosphere, leading to depth-dependent extension involving removal of lower crust, a process not considered in our earlier research¹⁴. The weak lower crust (Fig. 3) acts as the horizontal decoupling layer (Fig. 3, Supplementary Fig. 6c and Supplementary Movie 2), thereby allowing differential necking of lithospheric layers above and below. Model II-A evolves in two phases. The first comprises early stretching of a wide region of the upper and mid-crust matched by concomitant localized necking of the underlying mantle lithosphere. The small but regional stretching and thinning of the crust dies out with increasing distance into the continents. Phase one ends in rupture of the mantle lithosphere (Fig. 3a). During the second phase, the mantle lithosphere is advected away without further deformation by the plate motion, leaving the extending crust in contact with upwelled hot asthenosphere, which cools to form underplated 'oceanic' lithosphere (Fig. 3b). A characteristic of this model is that some hot, weak lower crust is extruded towards the rift axis as a pressure-driven channel flow (Fig. 3b). This material underplates regions of localized upper crustal extension (core complexes) and also exhumed at the rift axis¹⁵ (Fig. 3e). Most importantly, lower-crustal decoupling facilitates protracted crustal extension, leading to delayed crustal breakup, after forming an ultrawide margin.

Although central South Atlantic margins vary in style, we can see that some ultrawide ones have characteristics like those of model II-A (Fig. 3, and Supplementary Figs 2 and 3). Similar characteristics include ultrawide strongly attenuated crust (A), a wide region of sag basin subsidence (D), sag basin subsidence even in areas where there is little evidence of upper crustal extension, large post-rift subsidence consistent with model subsidence owing to cooling of hot oceanic lithosphere (E), and small to no syn-rift flank uplifts (F). We are not, however, aware of observational evidence for or against extrusive flow towards the rift axis.

A crucial but enigmatic observation from many of the ultrawide central South Atlantic margins is the late syn-rift/early post-rift lacustrine and shallow marine conditions of the sag basins^{16,17}, which persisted during salt deposition. This requires a syn-rift isostatic balance with relatively low-density subcrustal material, material that is more buoyant than upwelled asthenosphere. We propose that this material is hot depleted lower cratonic lithosphere. Model II-C (where C indicates cratonic inflow) demonstrates the mechanism (Fig. 3 and Supplementary Movie 3). It is a variant of model II-A in which rifting is adjacent to a craton (for example, the Congo craton in the central South Atlantic Ocean and the Pilbara craton in Exmouth Plateau) (Supplementary Figs 4 and 6b). The buoyant lower cratonic lithosphere preferentially flows into the subcrustal rift zone as the plates separate (Fig. 3c). In model II-C the underplating material has properties intermediate between highly depleted, high-viscosity craton mantle and the asthenosphere. Such properties are considered appropriate for partly depleted lower craton¹⁸⁻²⁰ with a parsimonious estimate of the depletion (compositional density anomaly (15 kg m^{-3} less than the density of the asthenosphere) and viscosity a factor $f = 3$ larger).

Model II-C (Fig. 3f) accounts for the following observations: (1) there is a lithospheric zone with high seismic shear velocity interpreted as 'continental material' under the ocean outboard of west Africa and restricted to the central South Atlantic²¹⁻²³; (2) this lithospheric zone is connected to the Congo craton^{21,24} (Supplementary Fig. 4); (3) there is evidence of continental crust and mantle lithosphere contamination in mantle magmas from this region²⁵⁻²⁷; (4) this part of the South Atlantic margin is relatively magma-poor, because cratonic inflow will suppress asthenospheric decompression melting, and (5) there is similar evidence from the Exmouth plateau^{24,28,29}. Additional evidence for cratonic mantle lithosphere beneath the Atlantic Ocean has recently been presented³⁰ and explained by syn-rift listric faulting and detachment of the continental crust, leaving pre-existing cratonic mantle lithosphere jutting out below the oceans. This explanation is probably incompatible with reconstruction of the initial fit of the continents, and

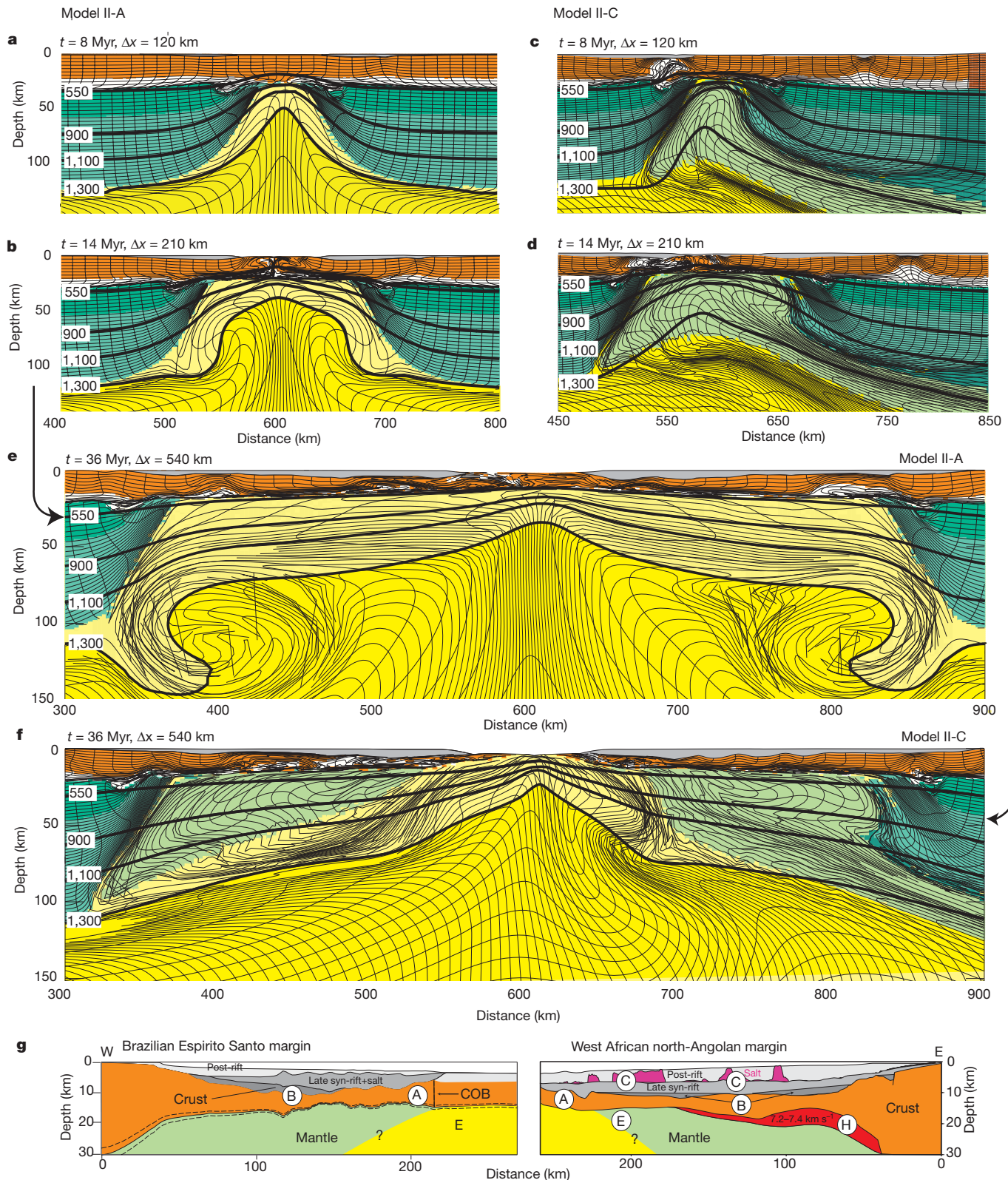


Figure 3 | Type II margins. a, b and e, Model II-A. c, d and f, Model II-C. g, Interpreted observations from central South Atlantic margins^{5,6,16,17,21–23}. Model style and colouring as in Fig. 2. Also shown are salt (magenta), early (dark grey) and late (medium grey) syn-rift sediments, possible magmatic

underplate (red), craton lower-mantle lithosphere (light green) and craton crust (brown). See Supplementary Information for model II-C explanation and Supplementary movies 2 and 3).

results in enhanced syn-rift subsidence of the margins, not reduced subsidence. We therefore prefer the lateral flow explanation.

Models like model II-C produce similar underplating flows for a range of parameter values, such that as f increases (for example, $f = 5$)

the trade-off requires a larger depletion density anomaly (-30 kg m^{-3}) to drive the flow. The margin bathymetry also depends on the density and thickness of the cratonic underplate and this could account for enigmatic shallow syn-rift bathymetry. Calculating the local isostatic

balance shows that underplating by a 60-km-thick cratonic layer with depletion density anomaly of -15 kg m^{-3} (compare with model II-C) reduces the syn-rift water-filled or vacuum-filled subsidence at any location by 391 m (water) or 273 m (vacuum) with respect to a case with no cratonic underplate (compare with model II-A). Both models II-A and II-C predict shallow syn-rift bathymetry even at the rift axis (Supplementary Movies 2 and 3 and Supplementary section 4) as a consequence of delayed crustal thinning combined with buoyant asthenospheric or cratonic underplate. In this case, cratonic underplate reduces the bathymetry only moderately. However, underplate with a depletion density anomaly of -50 kg m^{-3} reduces the water-filled or vacuum-filled subsidence of any location on the margin that has this underplate by 1,304 m (water) or 910 m (vacuum), causing the margin, including the rift axis, to remain subaerial for the first 20 million years of rifting (Supplementary section 4).

Exhumation of continental mantle lithosphere is not predicted for type II-A margins because most of the margin is underlain by upwelled asthenosphere (model II-A, Fig. 3e). However, exhumation is possible in type II-C margins if cratonic mantle lithosphere continues to flow to the rift axis and is exposed as the crust ruptures. Seismic reflection images from parts of the central South Atlantic are consistent with mantle exhumation (G. Karner, personal communication), but it is not known whether the mantle is continental or oceanic. No such lateral flow occurs for Phanerozoic mantle lithosphere in our models (see Fig. 3e), but is possible were the lower-mantle lithosphere to be depleted.

We propose that type I and type II margins are a direct consequence of their respective lithospheric rheological properties, which lead to contrasting styles of depth-dependent extension. These styles can be readily understood conceptually (Fig. 4) by considering the extension and necking of a laminate.

Type I lithosphere is a strongly bonded, frictional-plastic to viscous (approximately brittle–ductile) laminate that is weak only near its base (Fig. 4a). During extension, the upper lithosphere fails by faulting such that this layer ruptures while the lower lithosphere is still necking viscously (Fig. 4b, d). This style is achieved in model I by early excision of the lower crust and uppermost mantle, which places upper and mid-crust allochthons in contact with mid- and lower-mantle lithosphere as it is stretched, withdrawn, exhumed and exposed (Fig. 2d–f). The final step is rupture of the mantle lithosphere and onset of seafloor spreading. An essential requirement is depth-dependent extension in which the upper layer ruptures before the lower layer. This explains the observed exhumation of continental mid-mantle lithosphere⁹.

Type II lithosphere is a sandwich with a weak lower-crust viscous filling between two stronger layers (Fig. 4a). During extension the upper lithosphere decouples from the lower lithosphere over a wide region. The lower lithosphere necks in a similar manner to the type I laminate (Fig. 4b, c). However, when it breaks up, the upper lithosphere has thinned by only a minor amount because its extension is distributed across such a wide region (Fig. 4c, e). Rupture of the crust occurs much later, after it has been stretched like taffy to form a wide, thin layer bridging the severed lower-lithosphere conjugates (Fig. 4e). This model behaviour is not strongly influenced by cratonic underplating (Figs 3 and 4).

In conclusion, the characteristics of type I and II rifted margins, as shown in Fig. 1 and not explained by uniform lithospheric extension^{1,31}, can be explained by depth-dependent extension, as demonstrated by the models. The concept of different times of breakup (here used to imply ‘rupture’, not ‘rupture and accretion of oceanic crust’) of lower and upper lithosphere follows directly from the highly contrasting necking length scales of the upper lithosphere for the two cases. There are therefore two breakup events and accretion of oceanic crust occurs only after the second. In type I, upper lithosphere (crust) breakup occurs first, while the lower lithosphere is still necking, whereas in type II the reverse occurs. Uniform extension¹ corresponds to the special case in which both layers extend in the same way and undergo breakup at the same time. In addition, the enigmatic late syn-rift development of unfaulted, shallow-water sag basins at type II margins^{16,17} can be explained. Crustal extension, which

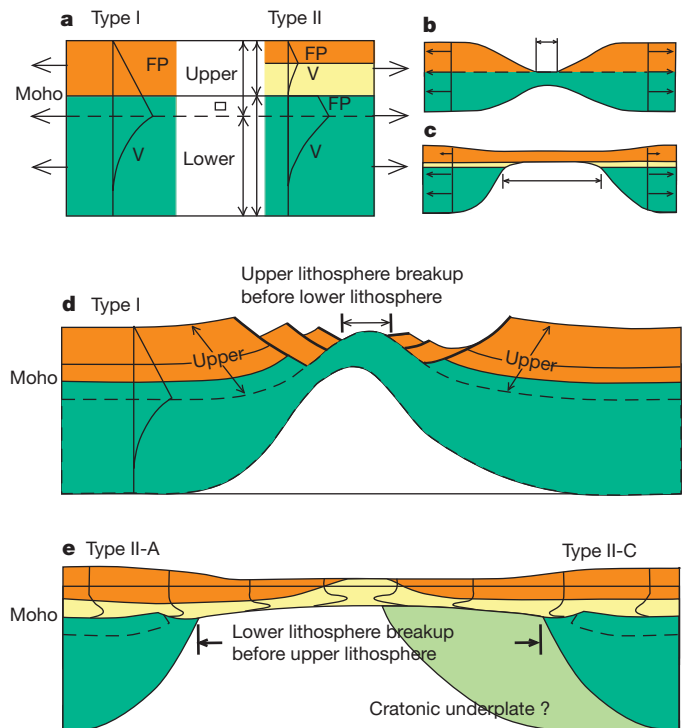


Figure 4 | Contrasting depth-dependent extension of type I and type II lithospheric laminates. **a**, Properties of the two types. FP, frictional-plastic; V, viscous; pale yellow, low-viscosity crust. **b**, **c**, Conceptual laminate necking styles showing respective early breakup of crust and mantle lithosphere. Adding isostasy gives: **d**, Type I. **e**, Type II-A and type II-C, with asthenospheric (left) and cratonic (right) underplate. We note that the region of the embedded cratonic lithosphere is to the right of panel **e**. Shown are strong FP crust (orange), low-viscosity crust (pale yellow), mantle lithosphere (green) and cratonic underplate (light green).

diachronously migrates across the margin towards the rift axis (models II-A and II-C), leaves late syn-rift sediments unfaulted, and subsidence of the margin is reduced owing to low-density cratonic underplate. The models and associated concepts are consistent with characteristics of these margins and their associated sedimentary basins. The pivotal differences from uniform extension¹ are the differential necking of the lithosphere, leading to a two-layer, two-stage breakup, and the possibility of cratonic underplate in addition to asthenospheric underplate.

METHODS SUMMARY

We use finite-element numerical models to calculate upper-mantle thermo-mechanically coupled, plane-strain, incompressible viscous–plastic creeping flows using an Arbitrary Lagrangian–Eulerian (ALE) method. When stress is below yield, the flow follows a power law, based on ‘wet’ quartzite and olivine laboratory measurements. Effective viscosity is specified by:

$$\eta = fA^{-1/n}(\dot{\gamma}_2')^{(1-n)/2n} \exp \left[\frac{Q+Vp}{nRT} \right] \quad (1)$$

where $\dot{\gamma}_2'$ is the second invariant of the deviatoric strain rate tensor $\frac{1}{2} \dot{\epsilon}_{ij} \dot{\epsilon}_{ij}$, n is the power-law exponent, A is the pre-exponential scaling factor, Q is the activation energy, V is the activation volume, p is the pressure, T is the absolute temperature and R is the universal gas constant. A , n , Q and V are derived from laboratory experiments and the parameter values are listed in Supplementary Table 1. The factor f is used to scale viscosities calculated from the reference quartz and olivine flow laws. This scaling produces strong and weak crust and reproduces the difference between ‘wet’ and ‘dry’ olivine (Methods and Supplementary Information, Supplementary Fig. 6c and Supplementary Table 1).

Frictional-plastic (Drucker–Prager) yielding occurs when:

$$\sigma_y = (J_2')^{1/2} = C \cos \phi_{\text{eff}} + p \sin \phi_{\text{eff}} \quad (2)$$

where $J_2' = \frac{1}{2} \sigma_{ij}' \sigma_{ij}'$ is the second invariant of the deviatoric stress, ϕ_{eff} is the effective internal angle of friction, C is cohesion, $\text{psin}(\phi_{\text{eff}}) = (p - p_i) \sin(\phi)$ and

p_f is the pore fluid pressure. This approximates frictional sliding in rocks, including pore-fluid pressure effects. Strain softening is introduced by a linear decrease of $\phi_{\text{eff}}(\epsilon)$ from 15° to 2° (Supplementary Fig. 6d and Supplementary Table 1). $\phi_{\text{eff}}(\epsilon) \approx 15^\circ$ corresponds to hydrostatic pore pressure.

In the thermal calculation the initial temperature for models I and II-A is laterally uniform, includes radioactive heat production in the crust ($A_R = 0.9 \mu\text{W m}^{-3}$) and basal heat flux ($q_m = 19.5 \text{ mW m}^{-2}$). In model II-C the craton has a reduced geothermal gradient. Cratonic lithosphere is compositionally depleted such that the upper and lower parts have $\rho_m(T_0) = 3,283 \text{ kg m}^{-3}$ (density anomaly, -17 kg m^{-3}) and $\rho_m(T_0) = 3,285 \text{ kg m}^{-3}$ (density anomaly, -15 kg m^{-3}). Mechanical and thermal systems are coupled and are solved sequentially at each time step.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 August 2010; accepted 2 March 2011.

- McKenzie, D. P. Some remarks on the development of sedimentary basins. *Earth Planet. Sci. Lett.* **40**, 25–31 (1978).
- Royden, L. & Keen, C. E. Rifting process and thermal evolution of the continental margin of eastern Canada determined from subsidence curves. *Earth Planet. Sci. Lett.* **51**, 343–361 (1980).
- Davis, M. & Kusznir, N. Depth-dependent lithospheric stretching at rifted continental margins. *Proc. NSF Rifted Margins Theoretical Institute* **1**, 92–136 (2004).
- Whitmarsh, R. B., Manatschal, G. & Minshull, T. A. Evolution of magma-poor continental margins from rifting to seafloor spreading. *Nature* **413**, 150–154 (2001).
- Moulin, M. *et al.* Geological constraints on the evolution of the Angolan margin based on reflection and refraction seismic data (ZaiAngo project). *Geophys. J. Int.* **162**, 793–810 (2005).
- Contrucci, I. *et al.* Deep structure of the West African continental margin (Congo, Zaire, Angola), between 5° S and 8° S, from reflection/refraction seismics and gravity data. *Geophys. J. Int.* **158**, 529–553 (2004).
- Peron-Pinvidic, G., Manatschal, G., Minshull, T. A. & Sawyer, D. S. Tectonosedimentary evolution of the deep Iberia-Newfoundland margins: evidence for a complex breakup history. *Tectonics* **26**, doi:10.1029/2006tc001970 (2007).
- Van Avendonk, H. J. A., Lavie, L. L., Shillington, D. J. & Manatschal, G. Extension of continental crust at the margin of the eastern Grand Banks, Newfoundland. *Tectonophysics* **468**, 131–148 (2009).
- Peron-Pinvidic, G. & Manatschal, G. The final rifting evolution at deep magma-poor passive margins from Iberia-Newfoundland: a new point of view. *Int. J. Earth Sci.* **98**, 1581–1597 (2009).
- Huismans, R. S. & Beaumont, C. Roles of lithospheric strain softening and heterogeneity in determining the geometry of rifts and continental margins. *Geol. Soc. Lond. Spec. Publ.* **282**, 111–138 (2007).
- Karner, G. D. & Driscoll, N. W. Style, timing and distribution of tectonic deformation across the Exmouth Plateau, northwest Australia, determined from stratal architecture and quantitative basin modelling. *Geol. Soc. Lond. Spec. Publ.* **164**, 271–311 (1999).
- Wernicke, B. Uniform-sense normal simple shear of the continental lithosphere. *Can. J. Earth Sci.* **22**, 108–125 (1985).
- Reston, T. Extension discrepancy at North Atlantic nonvolcanic rifted margins: depth-dependent stretching or unrecognized faulting? *Geology* **35**, 367–370 doi:10.1130/G23213a.1 (2007).
- Huismans, R. S. & Beaumont, C. Complex rifted continental margins explained by dynamical models of depth-dependent lithospheric extension. *Geology* **36**, 163–166 (2008).
- Aslanian, D. *et al.* Brazilian and African passive margins of the Central Segment of the South Atlantic Ocean: kinematic constraints. *Tectonophysics* **468**, 98–112 (2009).
- Karner, G. D. & Gamboa, L.A.P. Timing and origin of the South Atlantic pre-salt basins and their capping evaporites. *Geol. Soc. Lond. Spec. Publ.* **285**, 15–35 (2007).
- Dupre, S., Bertotti, G. & Cloetingh, S. Tectonic history along the South Gabon Basin: anomalous early post-rift subsidence. *Mar. Petrol. Geol.* **24**, 151–172 (2007).
- Griffin, W. L., O'Reilly, S. Y., Afonso, J. C. & Begg, G. C. The composition and evolution of lithospheric mantle: a re-evaluation and its tectonic implications. *J. Petrol.* **50**, 1185–1204 (2009).
- King, S. D. Archean cratons and mantle dynamics. *Earth Planet. Sci. Lett.* **234**, 1–14 (2005).
- Sleep, N. H. Evolution of the continental lithosphere. *Annu. Rev. Earth Planet. Sci.* **33**, 369–393 (2005).
- Begg, G. C. *et al.* The lithospheric architecture of Africa: seismic tomography, mantle petrology, and tectonic evolution. *Geosphere* **5**, 23–50 (2009).
- Ritsema, J. & van Heijst, H. New seismic model of the upper mantle beneath Africa. *Geology* **28**, 63–66 (2000).
- Sebai, A., Stutzmann, E., Montagner, J. P., Sicilia, D. & Beucler, E. Anisotropic structure of the African upper mantle from Rayleigh and Love wave tomography. *Phys. Earth Planet. Inter.* **155**, 48–62 (2006).
- McKenzie, D. & Priestley, K. The influence of lithospheric thickness variations on continental evolution. *Lithos* **102**, 1–11 (2008).
- Halliday, A. N., Dickin, A. P., Fallick, A. E. & Fitton, J. G. Mantle dynamics—a Nd, Sr, Pb and O isotopic study of the Cameroon Line volcanic chain. *J. Petrol.* **29**, 181–211 (1988).
- Rankenburg, K., Lassiter, J. C. & Brey, G. The role of continental crust and lithospheric mantle in the genesis of Cameroon Volcanic Line lavas: constraints from isotopic variations in lavas and megacrysts from the Biu and Jos plateaux. *J. Petrol.* **46**, 169–190 (2005).
- Regelous, M., Niu, Y. L., Abouchami, W. & Castillo, P. R. Shallow origin for South Atlantic Dupal Anomaly from lower continental crust: geochemical evidence from the Mid-Atlantic Ridge at 26° S. *Lithos* **112**, 57–72 (2009).
- Simons, F. J., Zielhuis, A. & van der Hilst, R. D. The deep structure of the Australian continent from surface wave tomography. *Lithos* **48**, 17–43 (1999).
- Fishwick, S., Heintz, M., Kennett, B. L. N., Reading, A. M. & Yoshizawa, K. Steps in lithospheric thickness within eastern Australia, evidence from surface wave tomography. *Tectonics* **27**, doi:10.1029/2007tc002116 (2008).
- O'Reilly, S. Y., Zhang, M., Griffin, W. L., Begg, G. & Hronsky, J. Ultradeep continental roots and their oceanic remnants: a solution to the geochemical “mantle reservoir” problem? *Lithos* **211S**, 1043–1054 (2009).
- Fraser, S. I., Fraser, A. J., Lentini, M. R. & Gawthorpe, R. L. Return to rifts—the next wave: fresh insights into the petroleum geology of global rift basins. *Petrol. Geosci.* **13**, 99–104 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements R.H. acknowledges support of the Department of Earth Science, University of Bergen, Norway. C.B. acknowledges support of the Canada Research Chair in Geodynamics. We thank G. Karner for comments on the manuscript.

Author Contributions R.H. contributed the numerical models and data for type I and II margins. C.B. contributed ideas on the cratonic underplate. Both authors contributed to writing the manuscript and to developing the concepts.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.H. (ritske.huismans@geo.uib.no).

METHODS

We use an Arbitrary Lagrangian–Eulerian (ALE) finite-element method for the solution of thermo-mechanically coupled, plane-strain, incompressible viscous–plastic creeping flows^{32–34} to investigate extension of a layered lithosphere with frictional–plastic and thermally activated power-law viscous rheologies (Supplementary Fig. 6).

When the state of stress is below the frictional–plastic yield the flow is viscous and is specified by temperature-dependent nonlinear power-law rheologies based on laboratory measurements on ‘wet’ quartzite³⁵ and ‘wet’ and ‘dry’ olivine³⁶. The effective viscosity η in the power-law rheology is of the general form:

$$\eta = fA^{-1/n}(\dot{\gamma}_2')^{(1-n)/2n} \exp\left[\frac{Q+Vp}{nRT}\right] \quad (1)$$

where $\dot{\gamma}_2'$ is the second invariant of the deviatoric strain rate tensor ($\frac{1}{2}\dot{\epsilon}_{ij}'\dot{\epsilon}_{ij}'$), n is the power-law exponent, A is the pre-exponential scaling factor, f is a scaling factor representing viscous weakening or strengthening, Q is the activation energy, V is the activation volume, which makes the viscosity dependent on pressure p , T is the absolute temperature, R is the universal gas constant, and $\dot{\epsilon}_{ij}'$ is the strain rate tensor. A (converted from the laboratory strain geometry to the tensor invariant form), n , Q and V are derived from the laboratory experiments and the parameter values are listed in Supplementary Table 1. We note that setting $V = 0$ for the quartzite flow law does not lead to significant errors because the pressure in the crust is low.

The effective viscosity of quartz-dominated rocks is characterized by large uncertainties. Supplementary Fig. 5 plots predicted effective viscosity as a function of temperature for a range of wet quartz flow laws (based on Table 3 of ref. 35). The reference parameter values for wet quartz used here, listed in Supplementary Table 1 (labelled ‘GT’ in Supplementary Fig. 5) represent a moderately strong viscous mid- and lower crust. The models described here use values of the reference quartz flow law which are scaled by factor f . The scale values are designed to produce crust with strong and weak regions (Supplementary Fig. 6c). In model I, strong crust with no viscous flow is achieved by increasing $\eta_{\text{wet quartz}}$ with scale factor $f = 30$. This can be interpreted to represent crust controlled by the viscous flow of ‘dry’ feldspar. In models II-A and II-C, weak crust is achieved with a scale factor of $f = 0.02$, which can be interpreted as viscous flow controlled by a weaker quartz rheology (Supplementary Fig. 5), an effect of viscous strain weakening, or a combination of both. In model II-C strong cratonic crust is achieved with a scale factor $f = 100$.

Frictional–plastic yielding is modelled with a pressure-dependent Drucker–Prager yield criterion, which is equivalent to the Coulomb yield criterion for incompressible deformation in plane strain. Yielding occurs when:

$$\sigma_y = (J_2')^{1/2} = C \cos \phi_{\text{eff}} + p \sin \phi_{\text{eff}} \quad (2)$$

where $J_2' = \frac{1}{2}\sigma_{ij}'\sigma_{ij}'$ is the second invariant of the deviatoric stress, σ_{ij}' is the deviatoric stress tensor, ϕ_{eff} is the effective internal angle of friction,

$p \sin(\phi_{\text{eff}}) = (p - p_f) \sin(\phi)$ for pore fluid pressure p_f , and C is cohesion. With appropriate choices of C and ϕ_{eff} this yield criterion can approximate frictional sliding in rocks and the effect of pore-fluid pressures. Plastic flow is incompressible. Strain softening is introduced by a linear decrease of $\phi_{\text{eff}}(e)$ from 15° to 2° (Supplementary Fig. 6d and Supplementary Table 1). We note that $\phi_{\text{eff}}(e) \approx 15^\circ$ corresponds to the effective ϕ when the pore fluid pressure is approximately hydrostatic.

In addition to solving the equilibrium equations for viscous–plastic flows in two dimensions, we also solve for the thermal evolution of the model. The mechanical and thermal systems are coupled through the temperature dependence of viscosity and density and are solved sequentially during each model time step. The initial temperature field for models I and II-A is laterally uniform, and increases with depth from the surface, $T_0 = 0^\circ\text{C}$, to the base of the crust, $T_m = 550^\circ\text{C}$, following a stable geotherm for uniform crustal heat production, $A_R = 0.9 \mu\text{W m}^{-3}$, and a basal heat flux, $q_m = 19.5 \text{ mW m}^{-2}$. Model I and II geothermal gradients, $8.6^\circ\text{C km}^{-1}$ and $0.4^\circ\text{C km}^{-1}$ (adiabatic), are uniform in the mantle lithosphere and sub-lithospheric mantle. Adiabatic heating and cooling are taken into account. Thermal boundary conditions are specified (basal temperature, $1,520^\circ\text{C}$) and lateral boundaries are insulated. Thermal diffusivity κ is $\kappa/\rho c_p = 10^{-6} \text{ m}^2 \text{ s}^{-1}$. For model II-C, the initial temperature field in the reference lithosphere is the same as in models I and II-A. In the cratonic part of the model, temperature increases linearly from about $T = 480^\circ\text{C}$ at the Mohorovičić discontinuity (Moho) to $T = 1,380^\circ\text{C}$ at the base of the cratonic lithosphere at $z = 250 \text{ km}$. The steady-state geothermal gradient in the cratonic mantle lithosphere was lowered by increasing the thermal diffusivity to, respectively, $\kappa_{\text{cml}} = 2.24 \times 10^{-6} \text{ m}^2 \text{ s}^{-1}$ and $\kappa_{\text{slm}} = 21.5 \times 10^{-6} \text{ m}^2 \text{ s}^{-1}$, thereby achieving a temperature of $1,380^\circ\text{C}$ at the base of the cratonic lithosphere. Densities of crust and mantle at 0°C are, respectively, $\rho_{0c} = \rho_c(T_0) = 2,800 \text{ kg m}^{-3}$ and $\rho_{0m} = \rho_m(T_0) = 3,300 \text{ kg m}^{-3}$, and depend on temperature with a volume coefficient of thermal expansion $\alpha_T = 2 \times 10^{-5}$ per $^\circ\text{C}$, $\rho(T) = \rho_0[1 - \alpha_T(T - T_0)]$. The cratonic lithosphere is depleted such that the upper cratonic mantle has $\rho_m(T_0) = 3,283 \text{ kg m}^{-3}$ (a compositional density anomaly of -17 kg m^{-3}) and the lower cratonic mantle lithosphere has $\rho_m(T_0) = 3,285 \text{ kg m}^{-3}$ (a compositional density anomaly of -15 kg m^{-3}).

32. Fullsack, P. An arbitrary Lagrangian–Eulerian formulation for creeping flows and its application in tectonic models. *Geophys. J. Int.* **120**, 1–23 (1995).
33. Willett, S. D. Rheological dependence of extension in wedge models of convergent orogens. *Tectonophysics* **305**, 419–435 (1999).
34. Huisman, R. S. & Beaumont, C. Symmetric and asymmetric lithospheric extension: relative effects of frictional–plastic and viscous strain softening. *J. Geophys. Res.* **108**, doi:10.1029/2002jb002026 (2003).
35. Gleason, G. C. & Tullis, J. A flow law for dislocation creep of quartz aggregates determined with the molten salt cell. *Tectonophysics* **247**, 1–23 (1995).
36. Karato, S. I. & Wu, P. Rheology of the upper mantle; a synthesis. *Science* **260**, 771–778 (1993).

Evolved structure of language shows lineage-specific trends in word-order universals

Michael Dunn^{1,2}, Simon J. Greenhill^{3,4}, Stephen C. Levinson^{1,2} & Russell D. Gray³

Languages vary widely but not without limit. The central goal of linguistics is to describe the diversity of human languages and explain the constraints on that diversity. Generative linguists following Chomsky have claimed that linguistic diversity must be constrained by innate parameters that are set as a child learns a language^{1,2}. In contrast, other linguists following Greenberg have claimed that there are statistical tendencies for co-occurrence of traits reflecting universal systems biases^{3–5}, rather than absolute constraints or parametric variation. Here we use computational phylogenetic methods to address the nature of constraints on linguistic diversity in an evolutionary framework⁶. First, contrary to the generative account of parameter setting, we show that the evolution of only a few word-order features of languages are strongly correlated. Second, contrary to the Greenbergian generalizations, we show that most observed functional dependencies between traits are lineage-specific rather than universal tendencies. These findings support the view that—at least with respect to word order—cultural evolution is the primary factor that determines linguistic structure, with the current state of a linguistic system shaping and constraining future states.

Human language is unique amongst animal communication systems not only for its structural complexity but also for its diversity at every level of structure and meaning. There are about 7,000 extant languages, some with just a dozen contrastive sounds, others with more than 100, some with complex patterns of word formation, others with simple words only, some with the verb at the beginning of the sentence, some in the middle, and some at the end. Understanding this diversity and the systematic constraints on it is the central goal of linguistics. The generative approach to linguistic variation has held that linguistic diversity can be explained by changes in parameter settings. Each of these parameters controls a number of specific linguistic traits. For example, the setting ‘heads first’ will cause a language both to place verbs before objects (‘kick the ball’), and prepositions before nouns (‘into the goal’)^{1,7}. According to this account, language change occurs when child learners simplify or regularize by choosing parameter settings other than those of the parental generation. Across a few generations such changes might work through a population, effecting language change across all the associated traits. Language change should therefore be relatively fast, and the traits set by one parameter must co-vary⁸.

In contrast, the statistical approach adopted by Greenbergian linguists samples languages to find empirically co-occurring traits. These co-occurring traits are expected to be statistical tendencies attributable to universal cognitive or systems biases. Among the most robust of these tendencies are the so-called “word-order universals”³ linking the order of elements in a clause. Dryer has tested these generalizations on a worldwide sample of 625 languages and finds evidence for some of these expected linkages between word orders⁹. According to Dryer’s reformulation of the word-order universals, dominant verb–object ordering correlates with prepositions, as well as relative clauses and genitives

after the noun, whereas dominant object–verb ordering predicts postpositions, relative clauses and genitives before the noun⁴. One general explanation for these observations is that languages tend to be consistent (‘harmonic’) in their order of the most important element or ‘head’ of a phrase relative to its ‘complement’ or ‘modifier’³, and so if the verb is first before its object, the adposition (here preposition) precedes the noun, while if the verb is last after its object, the adposition follows the noun (a ‘postposition’). Other functionally motivated explanations emphasize consistent direction of branching within the syntactic structure of a sentence⁹ or information structure and processing efficiency⁵.

To demonstrate that these correlations reflect underlying cognitive or systems biases, the languages must be sampled in a way that controls for features linked only by direct inheritance from a common ancestor¹⁰. However, efforts to obtain a statistically independent sample of languages confront several practical problems. First, our knowledge of language relationships is incomplete: specialists disagree about high-level groupings of languages and many languages are only tentatively assigned to language families. Second, a few large language families contain the bulk of global linguistic variation, making sampling purely from unrelated languages impractical. Some balance of related, unrelated and areally distributed languages has usually been aimed for in practice^{11,12}.

The approach we adopt here controls for shared inheritance by examining correlation in the evolution of traits within well-established family trees¹³. Drawing on the powerful methods developed in evolutionary biology, we can then track correlated changes during the historical processes of language evolution as languages split and diversify. Large language families, a problem for the sampling method described above, now become an essential resource, because they permit the identification of coupling between character state changes over long time periods. We selected four large language families for which quantitative phylogenies are available: Austronesian (with about 1,268 languages¹⁴ and a time depth of about 5,200 years¹⁵), Indo-European (about 449 languages¹⁴, time depth of about 8,700 years¹⁶), Bantu (about 668 or 522 for Narrow Bantu¹⁷, time depth about 4,000 years¹⁸) and Uto-Aztecan (about 61 languages¹⁹, time-depth about 5,000 years²⁰). Between them these language families encompass well over a third of the world’s approximately 7,000 languages. We focused our analyses on the ‘word-order universals’ because these are the most frequently cited exemplary candidates for strongly correlated linguistic features, with plausible motivations for interdependencies rooted in prominent formal and functional theories of grammar.

To test the extent of functional dependencies between word-order variables, we used a Bayesian phylogenetic method implemented in the software BayesTraits²¹. For eight word-order features we compared correlated and uncorrelated evolutionary models. Thus, for each pair of features, we calculated the likelihood that the observed states of the characters were the result of the two features evolving independently, and compared this to the likelihood that the observed states were the result of coupled evolutionary change. This likelihood calculation was

¹Max Planck Institute for Psycholinguistics, Post Office Box 310, 6500 AH Nijmegen, The Netherlands. ²Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Kapittelweg 29, 6525 EN Nijmegen, The Netherlands. ³Department of Psychology, University of Auckland, Auckland 1142, New Zealand. ⁴Computational Evolution Group, University of Auckland, Auckland 1142, New Zealand.

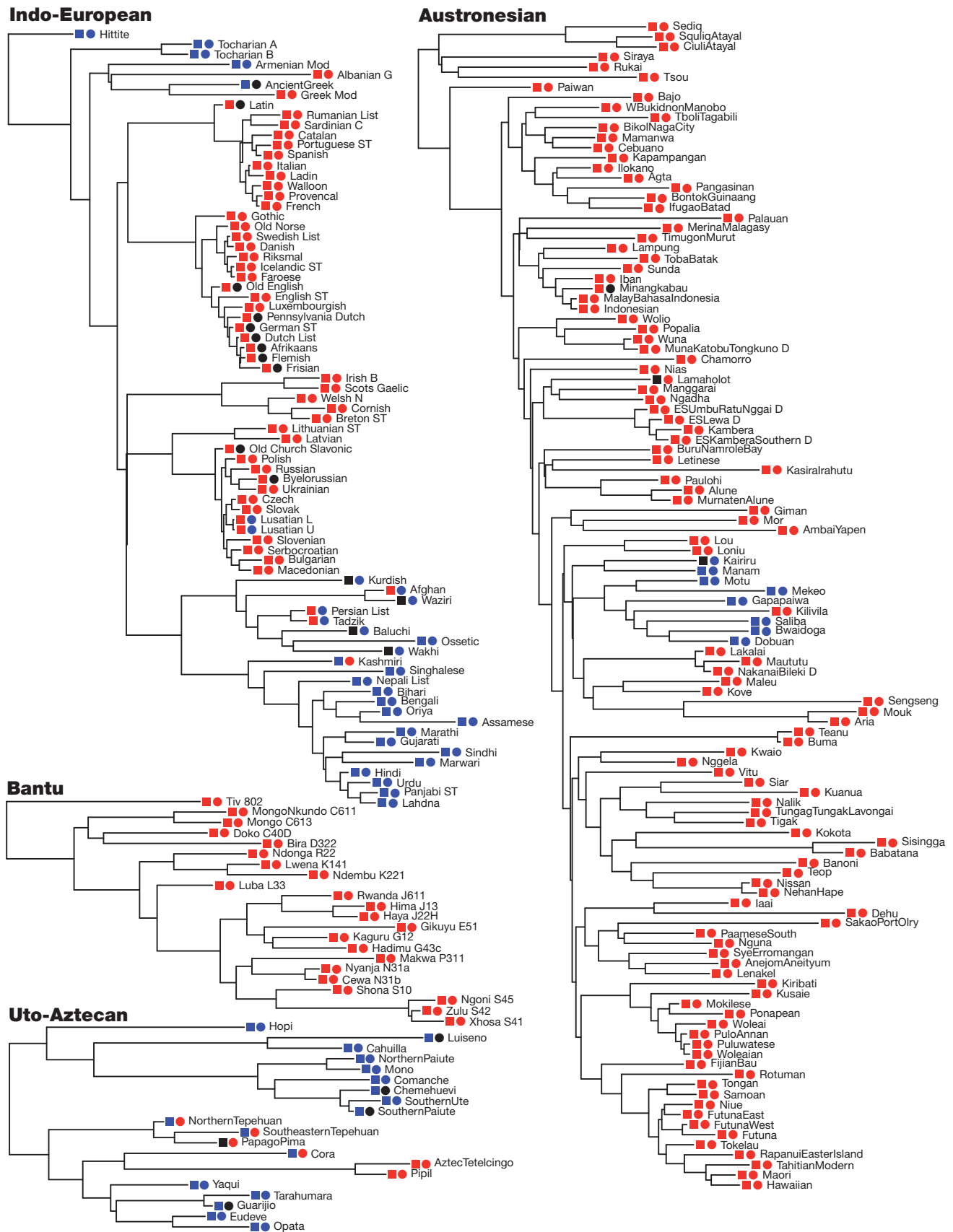


Figure 1 | Two word-order features plotted onto maximum clade credibility trees of the four language families. Squares represent order of adposition and noun; circles represent order of verb and object. The tree sample underlying this tree is generated from lexical data^{16,22}. Blue-blue indicates postposition,

object-verb. Red-red indicates preposition, verb-object. Red-blue indicates preposition, object-verb. Blue-red indicates postposition, verb-object. Black indicates polymorphic states.

conducted over a posterior probability distribution of phylogenetic trees constructed using basic vocabulary data from each of the language families: 79 Indo-European languages^{16,22}, 130 Austronesian languages^{15,23}, 66 Bantu languages²⁴ and 26 Uto-Aztecan languages (R. Ross & R.D.G., manuscript in preparation). Information on word-order typology was derived partly from the World Atlas of Language Structure database²⁵ and expanded with additional coding from grammatical descriptions (Supplementary Information section 1.3 and 2). As an illustration, the states of two of these features mapped against a summary of the posterior tree samples for all four language families are shown in Fig. 1. In this case, visual inspection shows that these characters appear to be linked in some families. However, the Bayesian phylogenetic approach allows us to assess this formally by quantifying the relative fits of dependent and independent models of character evolution across all trees in the posterior probability distribution. This method incorporates the uncertainty in the estimates of the tree topology, the rates of change and the branch lengths. The extent to which a dependent model of evolution provides a superior explanation of the variation of word-order features to an independent model is measured using Bayes factors (BF) calculated from the marginal likelihoods over the posterior tree distribution. $BF > 5$ are conventionally taken as strong evidence that the dependent model is preferred over the independent model^{13,26}.

The results of the BayesTraits analysis of correlated trait evolution are summarized in Fig. 2. These differ considerably from the expectations derived from both universal approaches. The Greenbergian approach suggests robust tendencies towards linkages due to intrinsic system biases, while the generative approach assumes these will be 'hard' systems constraints set by discrete choices over a small innate parameter set^{1,27}. Instead, our major finding is that, although there are linkages or dependencies between word-order characters within language families, these are largely lineage-specific, that is, they do not hold across language families in the way the two universals approaches predict.

Dryer's study of the Greenberg word-order universals⁴ across a world-wide sample of related and unrelated languages found a set of dependent word-order relations that show correlations with the order of verb and object, and another set of word-order relations that were independent of this. We extracted from his analyses two predictions of strong tendencies across all languages. First, all the word-order relations in the dependent set should be correlated: these are verb-object order, adposition-noun order, genitive-noun order, relative-clause-noun order. Second, no dependencies are expected between the dependent set and the independent set (including demonstrative-noun, numeral-noun, adjective-noun and subject-verb orders).

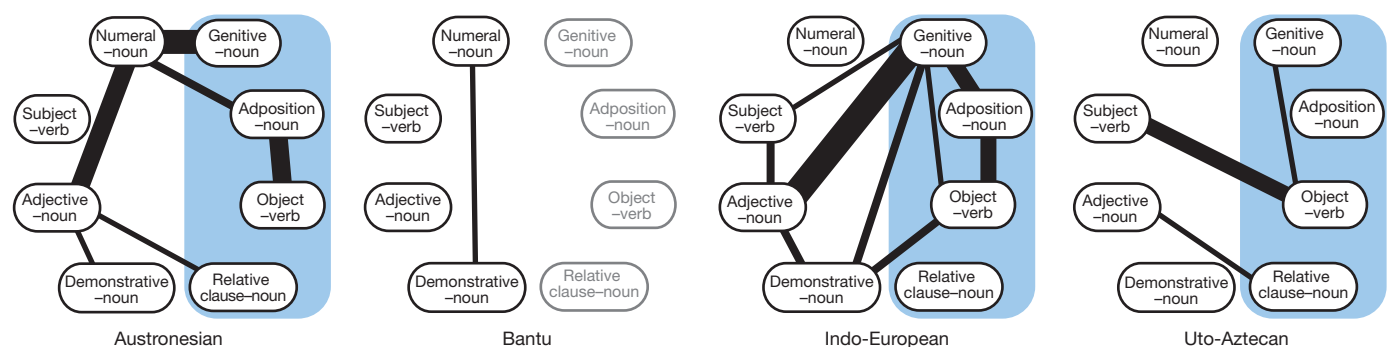


Figure 2 | Summary of evolutionary dependencies in word order for four language families. All pairs of characters where the phylogenetic analyses detect a strong dependency (defined as $BF \geq 5$) are shown with line width proportional to BF values (indicating a range from 5.01 to 21.23, see Supplementary Information section 3). In the case of the Bantu language family, four invariant features (indicated in grey) were excluded from the analyses. Following Dryer's reformulation of Greenberg's word-order universals, we expected dependencies between all the features in the blue

Contrary to the first expectation, we found no pairs of word-order features that were strongly dependent in all language families. Only two of these predicted dependencies were found in more than one language family: a dependency between adposition-noun and verb-object order was found in Austronesian and Indo-European, and a dependency between genitive-noun order and object-verb order was found in Indo-European and Uto-Aztecan.

Contrary to the second prediction, we found eight strong dependencies between members of the dependent set and members of the independent set, including two that occurred in two language families. The evolution of adjective-noun order and relative-clause-noun order is correlated within both Austronesian ($BF = 5.33$) and Uto-Aztecan ($BF = 5.02$), and the demonstrative-noun, object-verb features are correlated in Bantu ($BF = 5.24$) and Indo-European ($BF = 7.55$). Many dependencies are unique to just one language family; for example, only Uto-Aztecan shows strongly coupled ($BF = 13.57$) changes between subject and object ordering with respect to the verb, only Indo-European shows strongly coupled ($BF = 21.23$) changes between adjective and genitive ordering, and only Austronesian shows strongly coupled ($BF = 18.26$) changes between numeral-noun and genitive-noun orders. These family-specific linkages suggest that evolutionary processes of language diversification explore alternative ways to construct coherent language systems unfettered by tight universal constraints. They also demonstrate the power of phylogenetic methods to reveal structural linkages that could not be detected by cross-linguistic sampling.

The lineage-specificity of these dependencies is striking. There is a poor correspondence between dependencies across the families, and even where we find dependencies shared across language families, the phylogenetic analyses show family-specific evolutionary processes at work. Take, for example, the dependency between object-verb and adposition-noun orders shared by two of the language families. Examination of the transition probabilities between linked states reveals that different patterns of change are responsible for the observed linkage in each language family, as shown in Fig. 3. Here changes in the Austronesian family funnel evolving systems towards a single solution, while Indo-European shunts changes towards two solutions. Thus similarities in word-order dependencies may hide underlying differences in how these linkages come about, which once again reflect lineage-specific processes.

If the central goal of linguistic theory is to understand constraints on linguistic variation and language change, then the methods outlined here promise systematic insights of a kind only possible with the recent development of phylogenetic methods and large linguistic databases. As more large linguistic databases become available²⁸, the approach

shaded area. However, only two dependencies (object-verb order and adposition-noun order; and object-verb order and genitive-noun order) are found in more than one language family, and no dependencies were found involving relative-clause order and any of the other three features. Of the other thirteen strongly supported dependencies, nine were unexpected (no prediction was made about feature pairs outside the blue area). Most of these 19 dependencies occur in only one language family (three occur in two families, and one in three families).

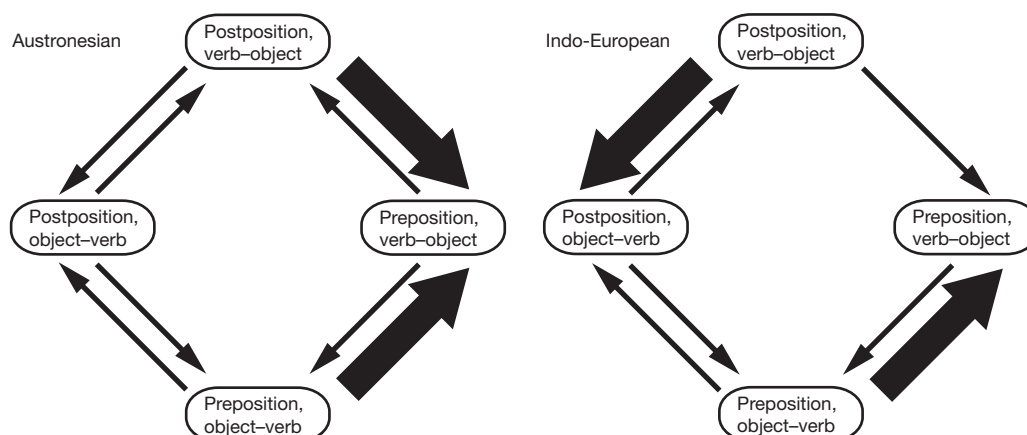


Figure 3 | The transition probabilities between states leading to object-verb and adposition-noun alignments in Austronesian and Indo-European. Data were taken from the model most frequently selected in the analyses; probability is indicated by line weight. The state pairs across the midline of each

figure (postposition, object-verb; and preposition, verb-object) are Greenberg's 'harmonic' or stable word orders. Nevertheless, each language family shows tendencies for specific directions and probabilities of state transitions.

developed here could be used to explore the dependency relationships between a wide range of linguistic features. Nearly all branches of linguistic theory have predicted such dependencies. Here we have examined the paradigm example (word-order universals) of the Greenbergian approach, taken also by the Chomskyan approach as "descriptive generalizations that should be derived from principles of UG [Universal Grammar]"^{1,27}. What the current analyses unexpectedly reveal is that systematic linkages of traits are likely to be the rare exception rather than the rule. Linguistic diversity does not seem to be tightly constrained by universal cognitive factors specialized for language²⁹. Instead, it is the product of cultural evolution, canalized by the systems that have evolved during diversification, so that future states lie in an evolutionary landscape with channels and basins of attraction that are specific to linguistic lineages.

Received 10 June 2009; accepted 8 February 2011.

Published online 13 April 2011.

1. Baker, M. *The Atoms of Language* (Basic Books, 2001).
2. Chomsky, N. *Lectures on Government and Binding* (Foris, 1981).
3. Greenberg, J. H. in *Universals of Grammar* (ed. Joseph H. Greenberg) 73–113 (MIT Press, 1963).
4. Dryer, M. in *Language Typology and Syntactic Description* Vol. I *Clause Structure* 2nd edn (ed. Shopen, T.) 61–131 (Cambridge University Press, 2007).
5. Hawkins, J. A. in *Language Universals* (eds Christiansen, M. H., Collins, C. & Edelman, S.) 54–78 (Oxford University Press, 2009).
6. Croft, W. Evolutionary linguistics. *Annu. Rev. Anthropol.* **37**, 219–234 (2008).
7. Chomsky, N. *Knowledge of Language: its Nature, Origin and Use* (Praeger, 1986).
8. Lightfoot, D. W. The child's trigger experience—degree-0 learnability. *Behav. Brain Sci.* **12**, 321–334 (1989).
9. Dryer, M. S. The Greenbergian word order correlations. *Language* **68**, 81–138 (1992).
10. Mace, R. & Pagel, M. The comparative method in anthropology. *Curr. Anthropol.* **35**, 549–564 (1994).
11. Bakker, P. in *Oxford Handbook of Linguistic Typology* (ed. Song, J. J.) (Oxford University Press, 2010).
12. Cysouw, M. in *Quantitative Linguistics: An International Handbook* (eds Altmann, G., Köhler, R. & Piotrowski, R.) 554–578 (Mouton de Gruyter, 2005).
13. Pagel, M., Meade, A. & Barker, D. Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* **53**, 673–684 (2004).
14. Gordon, R. G. J. *Ethnologue: Languages of the World* 15th edn (SIL International, 2005).
15. Gray, R. D., Drummond, A. J. & Greenhill, S. J. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* **323**, 479–483 (2009).

16. Gray, R. D. & Atkinson, Q. D. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* **426**, 435–439 (2003).
17. Guthrie, M. *Comparative Bantu* Vol. 2 (Gregg International, 1971).
18. Diamond, J. & Bellwood, P. Farmers and their languages: the first expansions. *Science* **300**, 597–603 (2003).
19. Campbell, L. *American Indian Languages: The Historical Linguistics of Native America* 133–138 (Oxford University Press, 1997).
20. Kemp, B. M. *et al.* Evaluating the farming/language dispersal hypothesis with genetic variation exhibited by populations in the Southwest and Mesoamerica. *Proc. Natl Acad. Sci. USA* **107**, 6759–6764 (2010).
21. Pagel, M. & Meade, A. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am. Nat.* **167**, 808–825 (2006).
22. Dyen, I., Kruskal, J. B. & Black, P. An Indo-European classification, a lexicostatistical experiment. *Trans. Am. Phil. Soc.* **82**, 1–132 (1992).
23. Greenhill, S. J., Blust, R. & Gray, R. D. The Austronesian basic vocabulary database: from bioinformatics to lexomics. *Evol. Bioinform.* **4**, 271–283 (2008).
24. Holden, C. J. Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. *Proc. R. Soc. Lond. B* **269**, 793–799 (2002).
25. Haspelmath, M., Dryer, M. S., Gil, D. & Comrie, B. *The World Atlas of Language Structures Online* (Max Planck Digital Library, 2008).
26. Raftery, A. Approximate Bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika* **83**, 251–266 (1996).
27. Cela-Conde, C. & Marty, G. Noam Chomsky's minimalist program and the philosophy of mind. An interview. *Syntax* **1**, 19–36 (1998).
28. Reesink, G., Singer, R. & Dunn, M. Explaining the linguistic diversity of Sahul using population models. *PLoS Biol.* **7**, e1000241 (2009).
29. Evans, N. & Levinson, S. C. The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* **32**, 429–492 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Liberman for comments on our initial results and F. Jordan and G. Reesink for comments on drafts of this paper. L. Campbell, J. Hill, W. Miller and R. Ross provided and coded the Uto-Aztecan lexical data.

Author Contributions R.D.G. and M.D. conceived and designed the study. S.J.G., R.D.G. and M.D. provided lexical data and phylogenetic trees. M.D. coded word-order data, and conducted the phylogenetic comparative analyses with S.J.G. All authors were involved in discussion and interpretation of the results. All authors contributed to the writing with S.C.L. and M.D. having leading roles; M.D., R.D.G. and S.J.G. produced the Supplementary Information.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.D. (michael.dunn@mpi.nl).

Body plan innovation in treehoppers through the evolution of an extra wing-like appendage

Benjamin Prud'homme¹, Caroline Minervino¹, Mélanie Hocine¹, Jessica D. Cande¹, Aïcha Aouane¹, Héloïse D. Dufour², Victoria A. Kassner² & Nicolas Gompel¹

Body plans, which characterize the anatomical organization of animal groups of high taxonomic rank¹, often evolve by the reduction or loss of appendages (limbs in vertebrates and legs and wings in insects, for example). In contrast, the addition of new features is extremely rare and is thought to be heavily constrained, although the nature of the constraints remains elusive^{2–4}. Here we show that the treehopper (Membracidae) 'helmet' is actually an appendage, a wing serial homologue on the first thoracic segment. This innovation in the insect body plan is an unprecedented situation in 250 Myr of insect evolution. We provide evidence suggesting that the helmet arose by escaping the ancestral repression of wing formation imparted by a member of the *Hox* gene family, which sculpts the number and pattern of appendages along the body axis^{5–8}. Moreover, we propose that the exceptional morphological diversification of the helmet was possible because, in contrast to the wings, it escaped the stringent functional requirements imposed by flight. This example illustrates how complex morphological structures can arise by the expression of ancestral developmental potentials and fuel the morphological diversification of an evolutionary lineage.

Treehoppers, a small group of hemipteran insects related to cicadas⁹, have evolved a peculiar morphological structure known as the helmet. It expands dorsally over most of the body length and has diversified to extremes within the family, conveying most of the treehoppers' shape diversity (Fig. 1). The various forms, colours and textures of the helmet may mimic natural elements ranging from thorns or seeds to animal droppings or aggressive ants^{10,11}. Without their helmets, treehoppers are very similar to cicadas (Supplementary Fig. 1b). The helmet is exclusively shared by all treehopper species, indicating that it appeared very early in the treehoppers' evolutionary lineage (Supplementary Fig. 1a). This evolutionary pattern prompted us to investigate how the helmet evolved.

The anatomical nature and evolutionary origin of the helmet remain controversial. Although most studies consider the helmet to be merely an expansion of the pronotum, that is, an enlarged dorsal face (tergite) of the first thoracic segment¹² (T1), it has been suggested¹³ that it could be a T1 appendage, a statement rejected by later workers^{12,14}. The key feature to discriminate between a simple outgrowth and an actual appendage is the presence of a jointed articulation, making the structure movable relative to the rest of the body. We found that the helmet has some elastic mobility, for instance in *Pubilia modesta*, one of the treehopper species we examined in this study (Supplementary Movie 1), suggesting that it is connected to the body through flexible attachments. Indeed, histological sections revealed that the helmet is bilaterally attached to the segment by a complex articulation (Fig. 2d–g). The attachment points consist of thin, non-sclerotized (that is, flexible) cuticle flanked by thicker, sclerotized cuticle (Fig. 2f). This configuration of flexible and hard cuticle (Fig. 2f, g, insets) defines cuticular joints that connect appendages to the body¹⁵, and is typically found at the attachment points of T2 and T3 wings (Fig. 2g). Because the helmet is

attached to T1 by jointed articulations, it follows that it is a T1 dorsal appendage, a situation completely unexpected in extant insects. The treehoppers' helmet is therefore distinct from the thoracic expansions that evolved in other insect lineages, for instance in horn beetles¹⁶ or in various other hemipterans (Supplementary Fig. 2a–c), which are cuticular projections and not articulated appendages. The conclusion that the helmet is a bona fide appendage does not exclude the hypothesis that, from an evolutionary perspective, the helmet initially arose from cuticular expansions. In this gradualist picture, the prothoracic outgrowths observed in some hemipterans might represent evolutionary forerunners of the treehoppers' helmet.

The presence of an extra dorsal appendage in treehoppers represented a rare opportunity to address how this type of body plan innovation emerged: either *de novo* or through the redeployment of an existing developmental program. Unlike most appendages, which are obviously paired, the helmet appears externally as a single structure both in adult and nymphal stages (Fig. 2b, e). To trace the developmental origin of the helmet, we sectioned first-instar nymphs and found that the helmet originates from two bilateral primordia, which later fuse along the dorsal midline (Fig. 2a, c and Supplementary Figs 3 and 4). The helmet is therefore a T1 dorsal appendage with a bilateral origin. Because the only known dorsal thoracic appendages in insects are wings (on T2 and T3), we explored the possibility that the helmet is

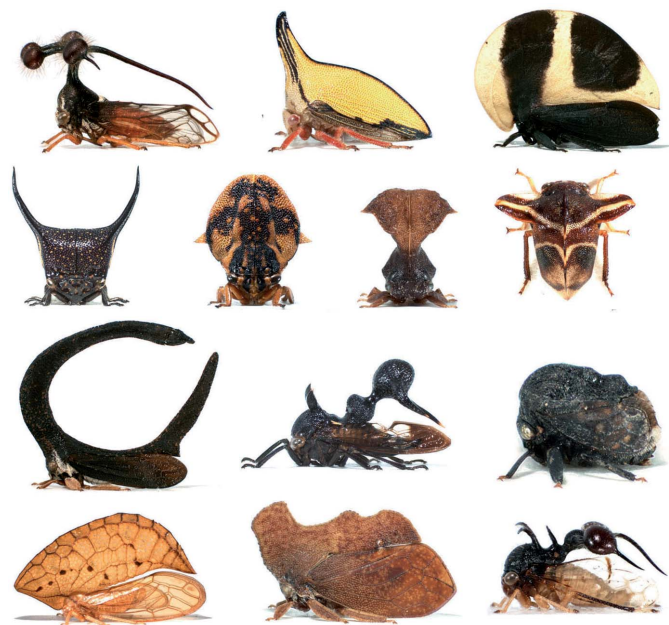


Figure 1 | Morphological diversity in treehoppers is conveyed by the helmet. Representative sample of neotropical treehopper (Membracidae) species (see Supplementary Table 1 for species names).

¹Institut de Biologie du Développement de Marseille-Luminy, CNRS UMR 6216, case 907, Parc scientifique de Luminy, 13288 Marseille cedex 9, France. ²University of Wisconsin and Howard Hughes Medical Institute, Bock Laboratories, 1525 Linden Drive, Madison, Wisconsin 53706, USA.

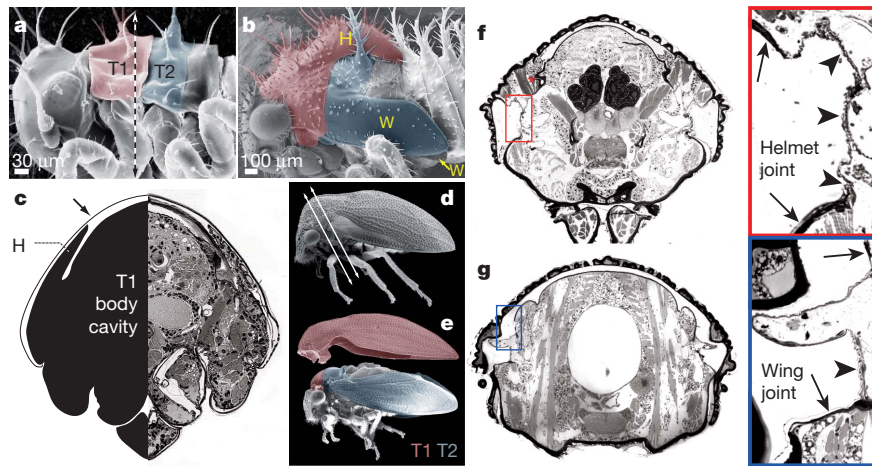


Figure 2 | The helmet is a T1 dorsal appendage with a bilateral origin.

a, b, Scanning electronic microscopy (SEM) images of nymphal stages 1 and 5 of *Publilia* sp., showing wings (W) on T2 (in blue) and the helmet (H) on T1 (in red). **c,** Sectioned (dotted arrowed line in **a**) first nymphal stage (right) and schematic (left); the external cuticle covers the helmet primordia (arrow).

a fused pair of wing serial homologues. Consistent with this notion, the wings and the helmet share several distinct morphological features: the helmet hinge consists of flexible, non-sclerotized cuticle (Fig. 2d, f) embedding small cuticular plates reminiscent of the pteralia that characterize the wing hinge region¹⁷ (Supplementary Fig. 5); both appendages consist of two layers of epithelial cells interconnected by large cuticular columns¹²; these layers unfold similarly on emergence, as any insect wing does (Supplementary Movie 2 and Supplementary Fig. 6); and a complex vein network covers both structures¹³ (Supplementary Fig. 1c). All together, these anatomical observations suggest that the helmet is a fused pair of wing serial homologues.

If the wings and the helmet are serial homologues, then their development must rely on a shared genetic program. We therefore searched for shared molecular signatures of wing and helmet development. A scant handful of transcription factors, including Nubbin¹⁸, mark wing developmental fate and allow for discrimination between wing and other appendage precursors¹⁹. We monitored the spatial deployment of Nubbin using a cross-reactive antibody¹⁹ and detected Nubbin expression during nymphal stages in the developing wings, as expected given its evolutionary conservation¹⁹ (Fig. 3a–d). Remarkably, Nubbin is found in the developing helmet also and its expression parallels that of the wings (Fig. 3b, c, e). Two other genes involved in the proximo-distal axis specification of appendages, *Distal-less* (*Dll*) and *homothorax* (*hth*), are also expressed in the developing helmet, and their distribution determines the helmet proximo-distal axis, from the hinge region to its posterior tip (Fig. 3f, g). These results suggest that the helmet and the wings share the same genetic program for their development, supporting the proposition that the treehopper's helmet is a T1 wing serial homologue.

The finding that treehoppers have evolved a T1 dorsal appendage is surprising in that all other extant winged insects have dorsal appendages restricted to T2 and T3 (ref. 20). This prompted the question of how the insect body plan has been modified in treehoppers. The fossil record indicates that the insect body plan progressively evolved some 350 Myr ago from one in which all segments bore wings or wing-like appendages to one in which the wings are confined to T2 and T3 (ref. 21). This transition was sculpted by *Hox* genes⁵, which evolved the ability to repress wing formation in the abdominal segments and T1. *Hox* gene repression of wing formation has been maintained for 250 Myr of insect evolution. In particular, *Sex combs reduced* (*Scr*) represses wing formation on T1 (Fig. 4e, left) through the repression of wing-growth and -patterning genes^{5,22–24}. For instance, when *Scr* is knocked down in *Tribolium*²⁵ (Coleoptera), ectopic wing primordia that express Nubbin

d, e, SEM images of intact (**d**) and dissected (**e**) *P. modesta* adults. **f, g,** Thick sections through a *P. modesta* adult thorax showing the helmet's articulation and the cuticular joints (boxes and insets) of helmet and wings (arrowheads point to thin, flexible cuticle, and arrows to thick cuticle). Muscles connect the helmet to the body (asterisk in **f**).

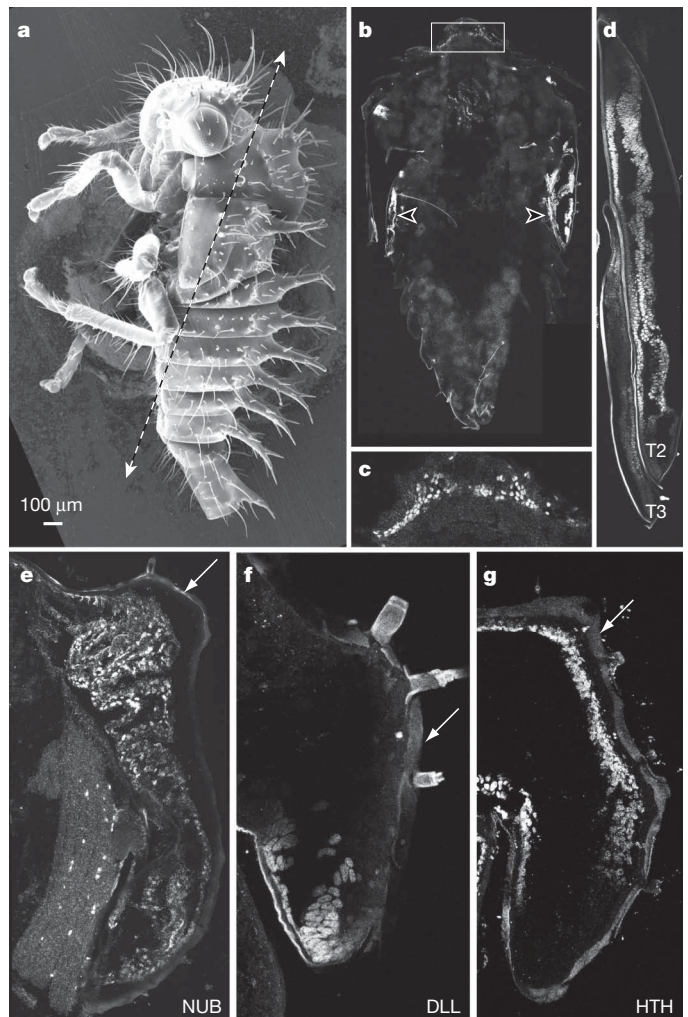


Figure 3 | Wing-patterning genes are expressed in the developing helmet.

a, *P. modesta* nymph (stage 4) showing the section plane of **b**. **b–e,** *P. modesta* stage-4 (**b, c**) and stage-5 (**d, e**) nymphs stained with an anti-NUB antibody. Sections reveal wing (**b**, arrowheads; **d**) and helmet (**c, e**) expressions. **f, g,** Sagittal sections stained with anti-DLL (**f**) and anti-HTH (**g**) antibodies; the bright outline surrounding the specimen is the auto-fluorescent cuticle (arrows in **e–g**). Specimens in **e–g** are at different nymphal stages.

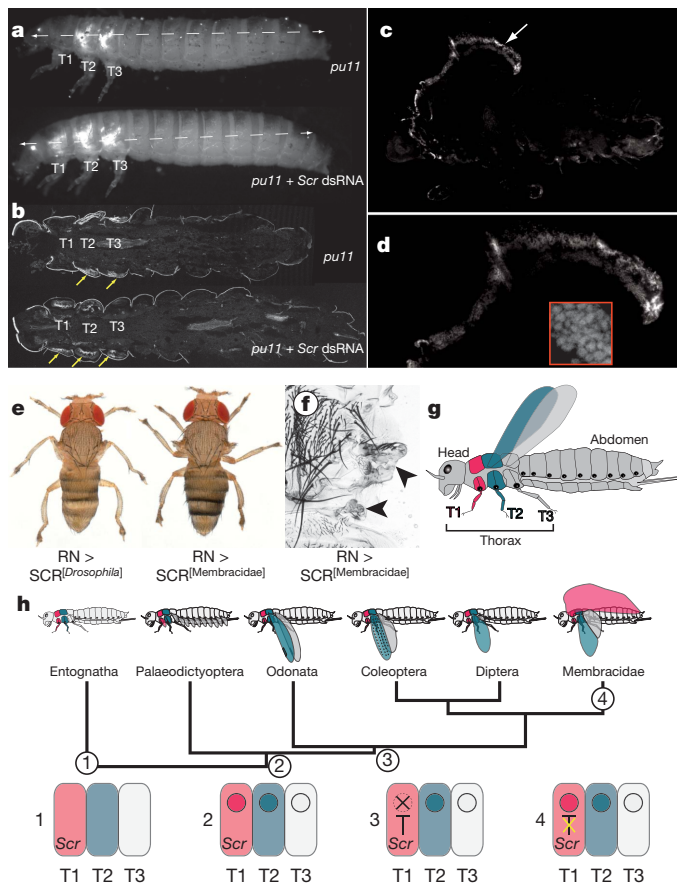


Figure 4 | *Scr* and the evolution of T1 appendages. **a, b,** Fifth-instar *Tribolium* *Cx^{ap1}* (ref. 22) larvae express the wing marker *pu11*-GFP (**a**, top panel), and Nubbin (arrows in **b**, top panel) in T2 and T3. When *Scr* is downregulated, ectopic wing primordia expressing Nubbin form on T1 (**a, b**, lower panels). **c, d,** Anti-*Scr* antibody staining on a *P. modesta* nymph (stage 5) sagittal section showing expression in T1, including the helmet (arrow in **c**; inset in **d** shows the nuclear distribution of the protein). **e, f,** Overexpression of *Drosophila* (**e**, left) and treehopper (**e**, right; **f**) *Scr* in fly imaginal discs abolishes wing and haltere formation (arrowheads in **f**). **g,** Generic body plan of a winged insect. **h,** Evolution of the regulatory link between *Scr* and the dorsal appendage development programme, from no link ancestrally (1, 2), to a repression (3) and the secondary loss in treehoppers (4).

form on T1 (Fig. 4a, b). This result shows that *Scr* prevents T1 wing formation through the repression of, at least, *nubbin* expression.

The expression of Nubbin in the developing treehoppers' helmet led us to propose that this structure evolved because *Scr* no longer exerted its ancestral repressive effect on wing formation, and we devised several possibilities that would account for this situation. First, *Scr* expression might be excluded from the helmet. We found, however, that *Scr* is expressed in the entire developing helmet (Fig. 4c, d), which is a priori incompatible with the T1-appendage-repressive function of *Scr* that is required until eclosion^{23,25}. Next we considered that in treehoppers *Scr* might have lost the ability to repress dorsal appendage development. We tested this possibility by ectopically expressing treehopper *Scr* in *Drosophila*. Ectopic expression of *Drosophila* *Scr*⁵ in fly wing and haltere precursors blocks their development (Fig. 4e, left). Similar ectopic expression of treehopper *Scr* results in identical phenotypes (Fig. 4e, right, and Fig. 4f). This result suggests that in treehoppers *Scr* is still capable of repressing T1 dorsal appendage development. All together, these results indicate that the evolution of the helmet is not due to a change in *Scr* expression or function, but rather to some genetic changes that occurred downstream of *Scr*. We propose that in treehoppers the wing developmental program, which involves *nubbin*, has

become unresponsive to *Scr* repression, possibly through selective regulatory changes downstream of *Scr* (Fig. 4g, h).

The distribution of wings along the body axis in insects seems particularly stable, as the only modifications in 250 Myr of evolution have been occasional losses or reductions²⁰. This body plan stability could be attributed to intrinsic developmental constraints that would prevent the evolution of extra appendages^{4,26}. Alternatively, it is conceivable that insects with extra sets of appendages do appear but are immediately counterselected. Identifying which type of constraint—developmental versus selective—limits the evolution of body plan has been a long-standing question³ that is difficult to address experimentally. Our results show that treehoppers have evolved a T1 dorsal appendage, thereby departing from the typical winged-insect body plan, by expressing a developmental potential that had been maintained under the repression of a *Hox* gene for 250 Myr. This argues that the constraint preventing extra dorsal appendage formation in insects is not developmental but rather selective. We submit that morphological innovations can arise from the deployment of existing but silenced developmental potentials, therefore requiring not so much the evolution of new genetic material but instead the expression of these potentials.

The breadth of morphological diversity in helmets that has evolved in less than 40 Myr (ref. 27 and C. Dietrich, personal communication) is unusual for an appendage. The pace of appendage evolution is generally slow, probably because of the strong selective pressure associated with their role in locomotion. This is particularly true for the wings²⁸, and we speculate that, initially alleviated from functional requirements, the recently evolved helmet was free to explore the morphological space through changes in its developmental program. A reminiscent pattern of appendage diversification on relaxed selection is observed for beetle elytra, which diverted from their primary flight function and have evolved all sorts of cuticular expansions, sculptures and glands²⁰ (Supplementary Fig. 7). More generally, these examples illustrate how a structure or an organ relieved from its original function (for instance by duplication or disuse), is “left to the free play of the various laws of growth”²⁹ and provides a new substrate for morphological diversification.

METHODS SUMMARY

Specimen collection. We collected *P. modesta* specimens in Wisconsin (USA).

Cloning and *Drosophila* genetics. UAS-*Scr* (*Drosophila* and treehopper) constructs were generated with standard cloning techniques and inserted at the same genomic position, preventing differences in transgenes activity due to position effects. *SCR* and Nubbin coding sequence alignments are shown in Supplementary Figs 8 and 9, respectively. *P. modesta* *Scr* and *nubbin* GenBank accession numbers are JF342360 and JF342361, respectively.

Immunocytochemistry. We used the following antibodies: anti-*SCR* (a gift from D. Andrews), anti-Nubbin (a gift from M. Averof), anti-DLL (a gift from S. Carroll) and anti-HTH (a gift from A. Salzberg).

For full details, see Supplementary Methods.

Received 30 December 2010; accepted 2 March 2011.

- Angelini, D. R. & Kaufman, T. C. Comparative developmental genetics and the evolution of arthropod body plans. *Annu. Rev. Genet.* **39**, 95–119 (2005).
- Maynard Smith, J. *et al.* Developmental constraints and evolution. *Q. Rev. Biol.* **60**, 265–287 (1985).
- Raff, R. A. *The Shape of Life: Genes, Development, and the Evolution of Animal Form* (Chicago Univ. Press, 1996).
- Riedl, R. A systems analytical approach to macroevolutionary phenomena. *Q. Rev. Biol.* **52**, 351–370 (1977).
- Carroll, S. B., Weatherbee, S. D. & Langeland, J. A. Homeotic genes and the regulation and evolution of insect wing number. *Nature* **375**, 58–61 (1995).
- Cohn, M. J. & Tickle, C. Developmental basis of limblessness and axial patterning in snakes. *Nature* **399**, 474–479 (1999).
- Di-Poi, N. *et al.* Changes in *Hox* genes' structure and function during the evolution of the squamate body plan. *Nature* **464**, 99–103 (2010).
- Pavlopoulos, A. *et al.* Probing the evolution of appendage specialization by *Hox* gene misexpression in an emerging model crustacean. *Proc. Natl Acad. Sci. USA* **106**, 13897–13902 (2009).
- Dietrich, C. H. Evolution of Cicadomorpha (Insecta, Hemiptera). *Denisia* **176**, 155–170 (2002).
- Fowler, W. W. in *Biologia Centrali Americana* Vol. II, Part 1 1–173 (Porter, 1894).

11. Poulton, E. B. in *Monograph of the Membracidae* (ed. Buckton, G. B.) 273–285 (Lovell Reeve, 1903).
12. Stegmann, U. E. An exaggerated trait in insects: the prothoracic skeleton of *Stictocephala bisonia* (Homoptera: Membracidae). *J. Morphol.* **238**, 157–178 (1998).
13. Richter, L. El apéndice pronotal en los Membracidos. *Lozania* **7**, 1–4 (1953).
14. Boulard, M. Le pronotum des Membracides: camouflage sélectionné ou orthogenèse hypertélique? *Bull. Mus. Hist. Nat. Paris (Zool.)* **83**, 145–165 (1973).
15. Snodgrass, R. E. *Principles of Insect Morphology* (McGraw-Hill, 1935).
16. Moczek, A. P., Rose, D., Sewell, W. & Kesselring, B. R. Conservation, innovation, and the evolution of horned beetle diversity. *Dev. Genes Evol.* **216**, 655–665 (2006).
17. Snodgrass, R. E. The thorax of insects and the articulation of the wings. *Proc. US Nat. Mus.* **36**, 511–595, pls 540–569 (1909).
18. Cifuentes, F. J. & Garcia-Bellido, A. Proximo-distal specification in the wing disc of *Drosophila* by the nubbin gene. *Proc. Natl Acad. Sci. USA* **94**, 11405–11410 (1997).
19. Averof, M. & Cohen, S. M. Evolutionary origin of insect wings from ancestral gills. *Nature* **385**, 627–630 (1997).
20. Grimaldi, D. & Engel, M. S. *Evolution of the Insects* (Cambridge Univ. Press, 2005).
21. Kukalová-Peck, J. Origin and evolution of insect wings and their relation to metamorphosis, as documented by the fossil record. *J. Morphol.* **156**, 53–125 (1978).
22. Beeman, R. W., Stuart, J. J., Haas, M. S. & Denell, R. E. Genetic analysis of the homeotic gene complex (HOM-C) in the beetle *Tribolium castaneum*. *Dev. Biol.* **133**, 196–209 (1989).
23. Chesebro, J., Hrycaj, S., Mahfooz, N. & Popadic, A. Diverging functions of Scr between embryonic and post-embryonic development in a hemimetabolous insect, *Oncopeltus fasciatus*. *Dev. Biol.* **329**, 142–151 (2009).
24. Rogers, B. T., Peterson, M. D. & Kaufman, T. C. Evolution of the insect body plan as revealed by the Sex combs reduced expression pattern. *Development* **124**, 149–157 (1997).
25. Tomoyasu, Y., Wheeler, S. R. & Denell, R. E. Ultrabithorax is required for membranous wing identity in the beetle *Tribolium castaneum*. *Nature* **433**, 643–647 (2005).
26. Brakefield, P. M. Evo-devo and constraints on selection. *Trends Ecol. Evol.* **21**, 362–368 (2006).
27. Poinar, G. & Poinar, R. *The Amber Forest: A Reconstruction of a Vanished World* (Princeton Univ. Press, 1999).
28. Comstock, J. H. *The Wings of Insects* (Comstock, 1918).
29. Darwin, C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (Murray, 1859).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to G. Moraguès for inspiring the project and the loan of specimens. We thank S. B. Carroll for discussions and support during the early stage of the project; M. Morgan and S. Morgan for access to their field property; J. P. Chauvin for assistance with SEM; S. McKamey and C. Dietrich for help with bibliography; D. Milo for discussions; and D. Andrews, M. Averof, S. B. Carroll, A. Kopp, A. Salzberg and Y. Tomoyasu for sharing reagents. We used FlyBase for information support. We also thank F. Leulier, T. Lecuit, M. Averof, C. Desplan and S. B. Carroll for comments on the manuscript. This work was supported by a EURYI award, a Human Frontier Science Program Career Development Award and the CNRS. J.D.C. was supported by a Human Frontier Science Program Long-term fellowship.

Author Contributions B.P. and N.G. conceived the project and designed the experiments; B.P., H.D.D., N.G. and V.A.K. collected *Publilia* specimens; B.P., C.M., J.D.C. and V.A.K. performed cloning; C.M., J.D.C., M.H. and N.G. did the immunostaining; and J.D.C., B.P. and N.G. carried out fly transgenesis and genetic experiments. A.A. made the histological sections, which were analysed by B.P. and N.G. N.G. and M.H. shot images and movies. N.G. made the anatomical dissections and observations. All authors participated in data analysis. B.P. and N.G. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.P. (benjamin.prudhomme@univmed.fr) or N.G. (nicolas.gompel@univmed.fr).

Functional specificity of local synaptic connections in neocortical networks

Ho Ko^{1*}, Sonja B. Hofer^{1*}, Bruno Pichler^{1†}, Katherine A. Buchanan¹, P. Jesper Sjöström¹ & Thomas D. Mrsic-Flogel¹

Neuronal connectivity is fundamental to information processing in the brain. Therefore, understanding the mechanisms of sensory processing requires uncovering how connection patterns between neurons relate to their function. On a coarse scale, long-range projections can preferentially link cortical regions with similar responses to sensory stimuli^{1–4}. But on the local scale, where dendrites and axons overlap substantially, the functional specificity of connections remains unknown. Here we determine synaptic connectivity between nearby layer 2/3 pyramidal neurons *in vitro*, the response properties of which were first characterized in mouse visual cortex *in vivo*. We found that connection probability was related to the similarity of visually driven neuronal activity. Neurons with the same preference for oriented stimuli connected at twice the rate of neurons with orthogonal orientation preferences. Neurons responding similarly to naturalistic stimuli formed connections at much higher rates than those with uncorrelated responses. Bidirectional synaptic connections were found more frequently between neuronal pairs with strongly correlated visual responses. Our results reveal the degree of functional specificity of local synaptic connections in the visual cortex, and point to the existence of fine-scale subnetworks dedicated to processing related sensory information.

Paired intracellular recordings in cortical slices indicate that synaptic connectivity between neighbouring neurons is heterogeneous and depends on factors such as cell type, electrophysiological properties and long-range targets^{5–10}. In fact, even within relatively homogenous groups of neurons, connectivity is not uniformly distributed^{5,6}. Although this non-random connectivity raises the possibility that functionally similar neurons form synaptically coupled subnetworks^{6,7}, the relationship between a neuron's synaptic partners and their functional properties in local cortical circuits has not been determined.

To elucidate this relationship, we developed an approach to relate connectivity to function in identified neurons of the layer 2/3 (L2/3) network in mouse visual cortex (V1), where neurons with diverse preferences for sensory stimuli are locally intermixed^{11,12}. In anaesthetized mice, the monocular region of V1 was bulk labelled with injections of the calcium indicator dye OGB-1 AM and the astrocyte marker SR101 (ref. 13) (see Methods). We first used *in vivo* two-photon imaging^{14,15} to sample spike-related somatic calcium signals from L2/3 neurons during presentation of drifting gratings and natural movie sequences (see Methods). We repeated this mapping at consecutive depths beneath the cortical surface to characterize visually evoked responses of all neurons within a cortical volume of approximately $285 \times 285 \times 90 \mu\text{m}^3$, starting at the upper border of L2/3 (Fig. 1a, b; depth range covered 60–120 μm). In this way, we obtained information about orientation/direction tuning and response correlation from a complete sample of L2/3 neurons (Fig. 1c, d).

We then identified the same OGB-1-filled neurons in acute slices (Figs 1e–h, 2a) by registering image stacks obtained *in vivo* and *in vitro* using affine transformation (see Methods and Supplementary Fig. 1),

and carried out simultaneous whole-cell patch-clamp recordings from up to four neighbouring L2/3 pyramidal neurons (mean distance \pm standard deviation (s.d.) = $25 \pm 9 \mu\text{m}$). Synaptic connectivity

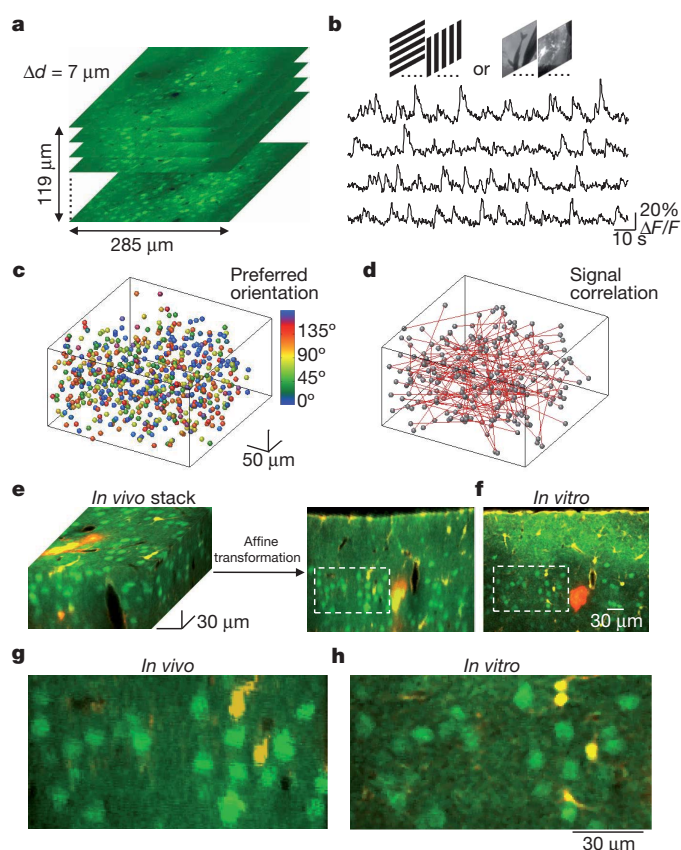


Figure 1 | Imaging functional properties of neurons *in vivo* and identifying the same neurons *in vitro*. **a**, Two-photon imaging was used to sample somatic calcium signals from a complete population of L2/3 neurons within a $285 \times 285 \times 119 \mu\text{m}^3$ volume. Imaging was carried out at $7 \mu\text{m}$ depth increments ($\Delta d = 7 \mu\text{m}$). Neurons were labelled with the calcium indicator dye OGB-1 AM (green) and the astrocyte marker SR101 (red). **b**, Example traces of calcium signals from four different cells in the imaged volume while presenting six trials of grating stimuli drifting in eight different directions. **c**, All orientation-selective cells in the volume were colour-coded according to preferred orientation and plotted as spheres. **d**, Signal correlations were computed from average responses to natural movies. Red lines represent strongly correlated neuronal pairs (signal correlation > 0.2). **e, f**, After imaging visually evoked calcium signals, a detailed image stack was obtained *in vivo*. The brain was sliced coronally and another stack of the same tissue was obtained *in vitro* (a single optical plane is shown in **f**). Affine transformation was used to align the *in vivo* to the *in vitro* stack, allowing precise matching of OGB-1-filled cells in the two stacks. **g, h**, Close-ups of the regions outlined with dashed lines in **e** and **f**, respectively.

¹Department of Neuroscience, Physiology and Pharmacology, University College London, 21 University Street, London WC1E 6DE, UK. [†]Present address: MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.

*These authors contributed equally to this work.

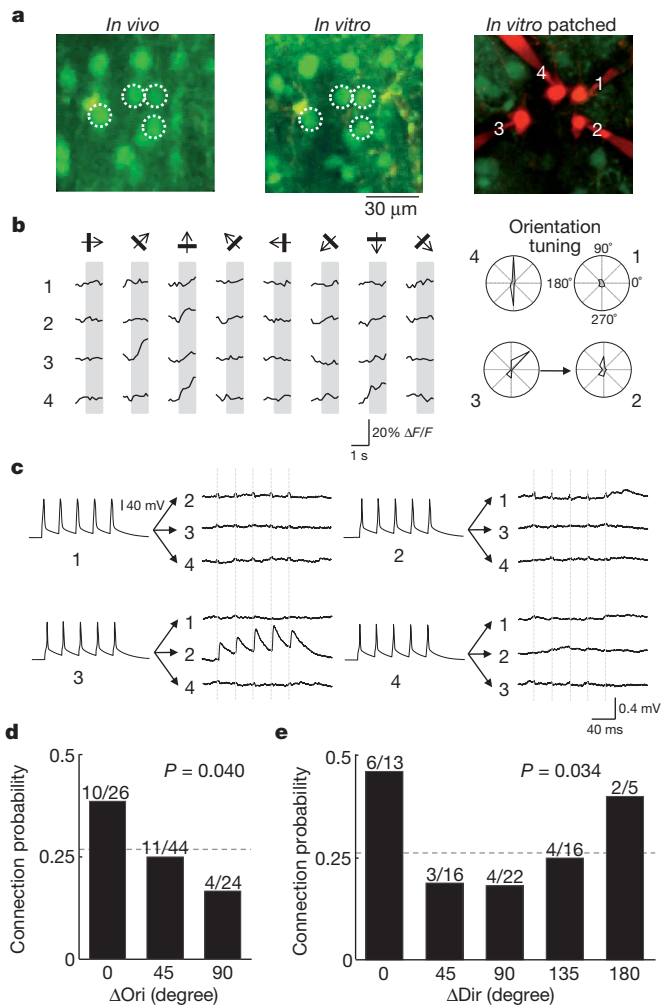


Figure 2 | Relating orientation and direction preference to connection probability among L2/3 pyramidal neurons. **a**, White circles denote the locations of *in vivo* to *in vitro* matched cells that were targeted for whole-cell recording and filled with Alexa 594. **b**, Left panel shows average calcium responses of the four cells to oriented drifting gratings. Right panel shows corresponding polar plots of inferred spike rate responses, normalized to the maximum response of cell 4. Three of the cells (cells 2, 3 and 4) were reliably responsive and orientation selective. Arrow shows a connection detected from cell 3 to cell 2. **c**, Membrane potential recordings from the four cells. Currents were injected into each cell in sequence, and from average traces of postsynaptic potentials an excitatory connection was found from cell 3 to cell 2. No other connections were found. Vertical dashed lines indicate timing of presynaptic spikes. In some traces, stimulation artefacts are visible that coincided exactly with presynaptic spikes and therefore could be clearly distinguished from EPSPs. **d**, Relationship between connection probability and difference in preferred orientation (ΔOri) among pairs in which both neurons were responsive to grating stimuli and were orientation selective ($\text{OSI} > 0.4$). There was a significant decreasing trend in connection probability as ΔOri increased ($P = 0.040$, Cochran–Armitage test). Dotted line indicates connection probability for all pairs included in this analysis (25/94, 0.27). The bins include difference in orientation values of 0 to 22.5° (0 degree bin), 22.5° to 67.5° (45 degree bin), and 67.5° to 90° (90 degree bin). **e**, Relationship between connection probability and difference in preferred direction (ΔDir) in the subset of neurons that were direction-selective ($\text{DSI} > 0.3$). The same decreasing trend with respect to ΔOri was detected ($P = 0.034$, Cochran–Armitage test). Dotted line indicates connection probability for all directionally selective pairs (19/72, 0.26). The bins include difference in orientation values of 0 to 22.5° (0 degree bin), 22.5° to 67.5° (45 degree bin), and so on.

between cells was assessed by evoking action potentials in each neuron in turn while simultaneously recording membrane potential in the other neurons. Monosynaptic connections appeared as spike-locked

postsynaptic potentials with millisecond latency (mean latency \pm s.e.m. = 1.69 ± 0.11 ms; see Figs 2c, 3b for sample traces). This approach allowed us to determine connectivity rates and patterns (unidirectional, bidirectional), and to relate these to cell functionality in the intact brain (Figs 1c, d, 2b, 3a).

The data set contained imaging experiments performed on 16 mice and whole-cell recordings from 126 L2/3 pyramidal cells, 116 of which could be matched to neurons functionally characterized *in vivo* (see Methods). The rate of connectivity was 0.19 (43 connections out of 222 potential connections assayed), in keeping with previous reports^{6,10}. Connection probability, synaptic strength and electrophysiological properties of OGB-1-labelled neurons were not significantly different to those recorded in slices from naive age-matched visual cortex that was not injected with OGB-1 AM (connectivity rate 0.18; 25 connected of 143 tested; Supplementary Fig. 2), indicating that dye loading, anaesthesia and prolonged exposure to infrared laser light during imaging *in vivo* did not alter these parameters.

We first examined how connectivity depended on orientation selectivity and on responsiveness to natural movies. Out of the 116 neurons, 77 were responsive to the natural movie, and 79 were orientation selective for grating stimuli (see Methods). Connection probability between orientation-tuned neurons was more than twofold higher than among non-selective and/or non-responsive cells (0.27; 25/94 versus 0.10; 3/31; $P = 0.050$, chi-squared test). The connectivity rate between neurons responsive to the natural movie was significantly higher than among cells non-responsive to the movie (0.28; 30/108 versus 0.04; 2/48; $P = 0.001$, chi-squared test). Taken together, these data indicate that reliably responsive and feature-selective neurons belong to more densely interconnected neocortical subnetworks.

We then related connection probability to neuronal preference for the angle and direction of drifting gratings (Fig. 2). For this analysis, we only included pairs in which both neurons were responsive (74/113), orientation selective (orientation selectivity index (OSI) > 0.4 ; 53/74), or direction selective (direction selectivity index (DSI) > 0.3 ; 41/53; see Methods and Supplementary Fig. 3a–c). Connectivity rate decreased with increasing difference in orientation preference ($P = 0.040$, Cochran–Armitage test for trend; Fig. 2d). For similarly tuned cells,

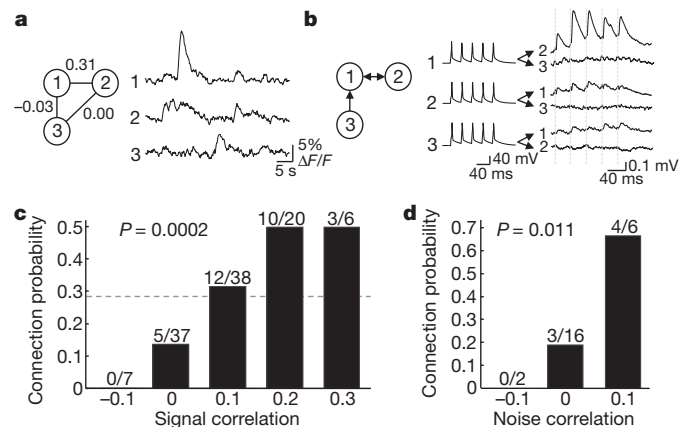


Figure 3 | Relationship between response correlation to natural movies and connection probability. **a**, An example of a triplet of neurons targeted for whole-cell recording *in vitro*, with associated *in vivo* calcium responses to the natural movie (average of six repetitions) and spike rate correlation values. Neuron 1 and 2 showed correlated firing (signal correlation = 0.31), whereas other pairs did not. **b**, Triple recordings from the same neurons reveal the pattern of connections: neurons 1 and 2 were bidirectionally connected, whereas neuron 3 provided input to neuron 1. Dashed lines indicate timing of presynaptic spikes. **c**, There was a significant increase in connection probability with increasing signal correlation to natural movies ($P = 0.0002$, Cochran–Armitage test). Dotted line indicates connection probability for all pairs included in this analysis (30/108, 0.28). **d**, Connection probability increased significantly with increase in noise correlation ($P = 0.011$, Cochran–Armitage test). Correlation values were binned, with ranges from -0.15 to -0.05 , from -0.05 to 0.05 , and so on.

connection probability was high (0.38; 10/26; difference in preferred orientation (ΔOri) $<22.5^\circ$), more than twofold higher than for cells with a large difference in orientation preference (0.17; 4/24; $\Delta\text{Ori} >67.5^\circ$). Thus, neuronal pairs similarly tuned for orientation were more likely to connect to each other, although a considerable connectivity rate was still observed between neurons tuned to dissimilar or orthogonal orientations. These results are consistent with the narrow suprathreshold yet broader subthreshold tuning for orientation and direction in mouse V1 neurons¹⁶. The same decrease of connection probability with increase in ΔOri was found for direction selective pairs ($P = 0.034$, Cochran–Armitage test; Fig. 2e), but these neurons only connected specifically with respect to orientation not preferred direction (Fig. 2e). These data indicate that directional preference is not conferred by biased local excitatory input, so other cell intrinsic or network mechanisms (for example, biased long range input, specific inhibition) may be needed to explain the emergence of direction selectivity. Varying the criteria for orientation or direction selectivity (OSI/DSI from 0.2 to 0.6) did not change the dependence of connectivity on difference in orientation/direction preference (Supplementary Fig. 3d, e), indicating that neurons that are broadly or sharply tuned both tend to connect preferentially to others with similar functional preference. In our data set we did not find evidence indicating that neurons with similar preferred orientations or directions are connected by stronger (excitatory postsynaptic potential (EPSP) amplitude) or more facilitating (paired-pulse ratio (PPR)) connections than neurons with different preferred orientations or directions (Supplementary Fig. 4a, b, d, e; also see Supplementary Fig. 5 for a sample pair with strong connections), although the sample size may not be adequate for ruling out any subtle trends.

The visual cortical circuit is constantly engaged in processing natural scenes, so statistical dependencies between neuronal activities in the presence of such stimuli may reflect connectivity. We therefore tested how network connectivity relates to the similarity of neuronal responses during the presentation of stimuli with natural spatiotemporal statistics (Fig. 3, see Methods). For each neuronal pair in which both neurons responded reliably to natural movies (56/113), we computed the time-varying firing-rate correlation of average responses (signal correlation) to repeated presentations of a 30 to 40-s-long natural movie sequence (Fig. 3a). On average, signal correlations were low (mean \pm s.d. = 0.08 ± 0.10). The probability of finding a connection between two neurons significantly increased with signal correlation to natural movies ($P = 0.0002$, Cochran–Armitage test; Fig. 3c). For pairs with close to zero or weakly negative signal correlation (<0.05), the connection probability was low (0.11, 5/44). In contrast, for neuronal pairs with stronger signal correlation (>0.15), the connection probability was more than fourfold higher (0.5; 13/26). Therefore, connectivity in mouse visual cortex is highly selective with respect to neuronal responses to natural movies. EPSP amplitude and PPR, however, were not found to change significantly with increase in signal correlation

(Supplementary Fig. 4c, f; also see Supplementary Fig. 5), although the sample size may not be large enough to rule out subtle trends.

Correlated variability in neuronal firing independent of a sensory stimulus is assumed to reflect neuronal connectivity in the network^{17–19}. Correlated fluctuations in neuronal firing may either be driven by common input or by recurrent synaptic connections, or both. For a subset of visually responsive neuronal pairs (12/56) that were imaged simultaneously *in vivo* (that is, on the same optical planes), we computed noise correlations (see Methods), which provide an indication of correlated response variability. Noise correlations were low (mean \pm s.d. = 0.02 ± 0.04). Despite the small sample size, connection probability was found to increase significantly with increase in noise correlation (Fig. 3d; $P = 0.011$, Cochran–Armitage test), indicating that recurrent connectivity may contribute to correlated fluctuations of neuronal firing.

We next compared how visual response similarity relates to connectivity motifs in the local network. Previous work indicates that bidirectional connections are overrepresented in a network of sparsely connected pyramidal neurons⁵. We found that the connectivity bias between neurons responding similarly to drifting gratings or to natural movies was further accentuated when investigating the distribution of unidirectionally or bidirectionally connected pairs (Fig. 4). We found a decreasing trend relating probability of bidirectional connections and difference in orientation preference (Fig. 4a, b; $P = 0.070$ for all orientation-selective pairs; $P = 0.036$ for direction-selective pairs, Cochran–Armitage test). Importantly, the monotonic fall-off in the incidence of bidirectional motifs was steeper than the overall decrease in probability of finding connected pairs as ΔOri increased (Fig. 4a, b). Similarly, the incidence of bidirectional connections increased sharply as signal correlation to natural movies increased ($P = 0.003$, Cochran–Armitage test; Fig. 4c), such that signal correlation was almost threefold higher for recurrently connected pairs than unconnected pairs (mean signal correlation of bidirectionally connected pairs \pm s.d. = 0.16 ± 0.07 versus 0.06 ± 0.10 for unconnected pairs; $P = 0.01$, rank sum test). As the probability of unidirectionally connected pairs did not show a monotonic trend with increase in response similarity (Fig. 4a–c; $P > 0.4$ for all conditions, Cochran–Armitage test), reciprocal connectivity reflects functional similarity better than does unidirectional connectivity.

In this study, we have characterized the functional specificity of local connections in mouse V1. Our results demonstrate that connectivity between neighbouring neurons ($<50\ \mu\text{m}$ apart) is not random, but specifically structured; visually driven neurons were more likely to connect to each other, and this probability increased with the degree of their response similarity. This relationship between connectivity and function was stronger when comparing responses to natural sensory input than for relatively artificial grating stimuli.

We have shown in mouse V1 that—although a given neuron receives input from nearby neurons preferring a wide range of stimulus

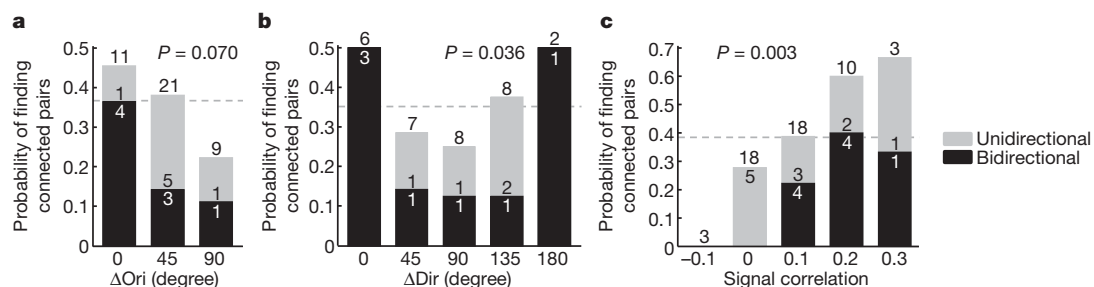


Figure 4 | Relationship between similarity of visual responses and probability of finding unidirectionally and bidirectionally connected pairs. **a**, Among orientation-selective neurons, the probability of finding connected pairs decreased as ΔOri increased. The fall-off in probability of finding bidirectionally connected pairs was steeper than the decrease in overall probability of finding connected pairs. A trend of decrease in probability of finding bidirectionally connected pairs was found ($P = 0.070$, Cochran–Armitage test). **b**, The same

observation holds in the subset of directionally selective pairs, and the probability of finding bidirectionally connected pairs decreased as ΔOri increased ($P = 0.036$, Cochran–Armitage test). **c**, The probability of finding bidirectionally connected pairs increased sharply as signal correlation to natural movies increased ($P = 0.003$, Cochran–Armitage test). Dotted lines indicate the probability of finding connected pairs from all pairs included in analysis (panel **a**: 15/41, 0.37; panel **b**: 11/31, 0.35; panel **c**: 20/52, 0.38).

orientations—more than twice as many connections are made between similarly tuned neurons as between disparately tuned cells. In keeping, subthreshold tuning in L2/3 pyramidal neurons in mouse V1 is broad but nonetheless biased towards the preferred orientation¹⁶. This is similar to the tuning of neurons in pinwheel centres of orientation maps in visual cortex of other species²⁰. In carnivores and primates, long-range horizontal projections in L2/3 (>500 µm) are biased towards cortical columns with similar orientation preference^{1–4}. Our results indicate that similar principles of connectivity apply at the level of local neocortical networks in the mouse—a species without columnar architecture—indicating that functionally biased connectivity may be a general feature of organization in the visual cortex. In the visual cortex this selective connection scheme may serve as a mechanism for the amplification of thalamic input and sharpening of tuning^{21,22} or for local contour integration²³.

Analysis of connectivity rate with respect to similarity of responses to natural movies revealed a marked degree of specificity of local connections (Fig. 3c, d). Connection probability increased sharply with increase in both signal and noise correlation to natural movies. Neurons with higher signal correlations to natural movies probably share similar receptive field structures, and may therefore be driven by common feed-forward input²⁴. Our results are therefore consistent with the finding that L2/3 pyramidal neurons form highly interconnected subnetworks sharing common input from layer 4 in slices of rat visual cortex⁶. Developmentally, this organization of lateral connections based on receptive field similarity may arise through activity-dependent synaptic plasticity, whereby neurons driven by common input develop stable bidirectional connections²⁵. Indeed, our data show that the majority of bidirectionally connected neurons had stronger signal correlations to natural movies and shared similar orientation preference. As individual neurons show variability in their responses to the same visual stimulus²⁶, recurrent excitation between similarly tuned neurons may reduce response variance, while introducing redundancy into the population code for robustness against errors²⁷.

Our results do not preclude the possibility that other factors—including inhibitory connections or synaptic strength—also contribute to functional specificity in the circuit. Because inhibition, in particular, may be important in determining the receptive field properties of neurons in V1 (ref. 28), it will be important to examine the extent to which inhibitory connections are functionally specific⁷.

Using a novel and relatively straightforward approach for *in vitro* mapping of synaptic connectivity among neurons that had been identified functionally *in vivo*, we found that neighbouring neurons with similar feature selectivity preferentially but not exclusively connected to each other in L2/3 of mouse V1. Together with other powerful approaches^{29,30}, our method can be used to uncover functional biases of connectivity between different cell types and cortical layers, and in other brain areas. This information will be critical for understanding the functional wiring of circuits mediating perception and behaviour.

METHODS SUMMARY

Anaesthetized C57Bl/6 mice between postnatal day 22 and 26 were injected with the calcium-sensitive dye Oregon Green Bapta-1 AM into monocular V1 as described previously¹¹ and *in vivo* two-photon calcium imaging^{14,15} was used to record responses of layer 2/3 neurons to eight different drifting square-wave gratings (0.035 cycles per degree, 2 cycles s⁻¹, 100% contrast) and natural movie sequences. Spike trains were inferred from calcium signals using a non-negative deconvolution method. Preferred orientation and direction, as well as OSI and DSI were calculated using Fourier-interpolated tuning curves. Pearson's correlation coefficient was used to obtain pair-wise response correlations, either from average responses to the stimulus (signal correlation) or from mean-subtracted responses (noise correlation). Small volumes of fluorescent microspheres were injected into the imaged region to facilitate identification of the region in the sliced brain. Coronal slices were cut after dissection of the brain, and whole-cell recordings from up to four cells simultaneously were carried out in the vicinity of the microsphere tract (identified by two-photon microscopy). The presence of synaptic connections was tested by evoking five spikes at 30 Hz in each cell, repeated

30–90 times. Connection probability is the number of detected connections over the total number of potential connections assayed. Probability of finding uni- or bidirectionally connected pairs was calculated as the number of uni- or bidirectionally connected pairs over the total number of pairs. To register *in vivo* and *in vitro* image stacks and to match the same neurons imaged *in vivo* and recorded from *in vitro*, three-dimensional image registration by affine transformation using custom-written MATLAB software was performed subsequent to the experiment. To relate connectivity to functional properties, the asymptotic Cochran–Armitage test for trend was used to test for significance.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 September 2010; accepted 1 February 2011.

Published online 10 April 2011.

1. Bosking, W. H., Zhang, Y., Schofield, B. & Fitzpatrick, D. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J. Neurosci.* **17**, 2112–2127 (1997).
2. Gilbert, C. D. & Wiesel, T. N. Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *J. Neurosci.* **9**, 2432–2442 (1989).
3. Roerig, B. & Kao, J. P. Organization of intracortical circuits in relation to direction preference maps in ferret visual cortex. *J. Neurosci.* **19**, RC44 (1999).
4. Weliky, M., Kandler, K., Fitzpatrick, D. & Katz, L. C. Patterns of excitation and inhibition evoked by horizontal connections in visual cortex share a common relationship to orientation columns. *Neuron* **15**, 541–552 (1995).
5. Song, S., Sjöström, P. J., Reigl, M., Nelson, S. & Chklovskii, D. B. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* **3**, e68 (2005).
6. Yoshimura, Y., Dantzker, J. L. M. & Callaway, E. M. Excitatory cortical neurons form fine-scale functional networks. *Nature* **433**, 868–873 (2005).
7. Yoshimura, Y. & Callaway, E. M. Fine-scale specificity of cortical networks depends on inhibitory cell type and connectivity. *Nature Neurosci.* **8**, 1552–1559 (2005).
8. Brown, S. P. & Hestrin, S. Intracortical circuits of pyramidal neurons reflect their long-range axonal targets. *Nature* **457**, 1133–1136 (2009).
9. Thomson, A. M., Bannister, A. P., Mercer, A. & Morris, O. T. Target and temporal pattern selection at neocortical synapses. *Phil. Trans. R. Soc. Lond. B* **357**, 1781–1791 (2002).
10. Holmgren, C., Harkany, T., Svennenfors, B. & Zilberter, Y. Pyramidal cell communication within local networks in layer 2/3 of rat neocortex. *J. Physiol. (Lond.)* **551**, 139–153 (2003).
11. Mrcic-Flogel, T. D. *et al.* Homeostatic regulation of eye-specific responses in visual cortex during ocular dominance plasticity. *Neuron* **54**, 961–972 (2007).
12. Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P. & Reid, R. C. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* **433**, 597–603 (2005).
13. Nimmerjahn, A., Kirchhoff, F., Kerr, J. N. D. & Helmchen, F. Sulforhodamine 101 as a specific marker of astroglia in the neocortex *in vivo*. *Nature Methods* **1**, 31–37 (2004).
14. Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning fluorescence microscopy. *Science* **248**, 73–76 (1990).
15. Stosiek, C., Garaschuk, O., Holthoff, K. & Konnerth, A. *In vivo* two-photon calcium imaging of neuronal networks. *Proc. Natl Acad. Sci. USA* **100**, 7319–7324 (2003).
16. Jia, H., Rochefort, N. L., Chen, X. & Konnerth, A. Dendritic organization of sensory input to cortical neurons *in vivo*. *Nature* **464**, 1307–1312 (2010).
17. Alonso, J. M. & Martinez, L. M. Functional connectivity between simple cells and complex cells in cat striate cortex. *Nature Neurosci.* **1**, 395–403 (1998).
18. Kohn, A. & Smith, M. A. Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J. Neurosci.* **25**, 3661–3673 (2005).
19. Aertsen, A. M., Gerstein, G. L., Habib, M. K. & Palm, G. Dynamics of neuronal firing correlation: modulation of “effective connectivity”. *J. Neurophysiol.* **61**, 900–917 (1989).
20. Mariño, J. *et al.* Invariant computations in local cortical networks with balanced excitation and inhibition. *Nature Neurosci.* **8**, 194–201 (2005).
21. Ben-Yishai, R., Bar-Or, R. L. & Sompolinsky, H. Theory of orientation tuning in visual cortex. *Proc. Natl Acad. Sci. USA* **92**, 3844–3848 (1995).
22. Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A. & Suarez, H. H. Recurrent excitation in neocortical circuits. *Science* **269**, 981–985 (1995).
23. Li, W., Piech, V. & Gilbert, C. D. Contour saliency in primary visual cortex. *Neuron* **50**, 951–962 (2006).
24. Alonso, J. M., Usrey, W. M. & Reid, R. C. Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex. *J. Neurosci.* **21**, 4002–4015 (2001).
25. Clopath, C., Büsing, L., Vasilaki, E. & Gerstner, W. Connectivity reflects coding: a model of voltage-based STDP with homeostasis. *Nature Neurosci.* **13**, 344–352 (2010).
26. Tolhurst, D. J., Movshon, J. A. & Dean, A. F. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.* **23**, 775–785 (1983).
27. Barlow, H. Redundancy reduction revisited. *Network* **12**, 241–253 (2001).
28. Alitto, H. J. & Dan, Y. Function of inhibition in visual cortical processing. *Curr. Opin. Neurobiol.* **20**, 340–346 (2010).
29. Marshel, J. H., Mori, T., Nielsen, K. J. & Callaway, E. M. Targeting single neuronal networks for gene expression and cell labeling *in vivo*. *Neuron* **67**, 562–574 (2010).

30. Denk, W. & Horstmann, H. Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol.* **2**, e329 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Margrie for discussions about the project and the manuscript, and J. Vogelstein for the spike inference algorithm. This work was supported by the Wellcome Trust (T.D.M.-F.), the European Research Council (T.D.M.-F.), the European Molecular Biology Organisation (S.B.H.), the Medical Research Council and FP7 grant #243914 (K.A.B., P.J.S.), the Overseas Research Students Award Scheme and UCL studentship (H.K.).

Author Contributions H.K. and S.B.H. performed experiments and data analysis. H.K. developed image registration software using preliminary data obtained by S.B.H. and K.A.B., and programs for data analysis. B.P. developed image acquisition software and the program for extracting calcium transients. P.J.S. designed electrophysiology setup and software for acquisition and analysis. B.P., H.K., S.B.H. and T.D.M.-F. built experimental setups. H.K. and T.D.M.-F. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.D.M.-F. (t.mrsic-flogel@ucl.ac.uk).

METHODS

Animals and surgical procedures. All experimental procedures were carried out in accordance with institutional animal welfare guidelines and were licensed by the UK Home Office. Experiments were performed on C57Bl/6 mice between postnatal day 22–26, when both intrinsic and visually driven cortical responses exhibit a relatively mature phenotype^{31,32}. Mice were initially anaesthetized with a mixture of Fentanyl (0.05 mg kg⁻¹), Midazolam (5.0 mg kg⁻¹), and Medetomidin (0.5 mg kg⁻¹). Light anaesthesia was maintained during recordings by isoflurane (0.3–0.5%) in a 60:40% mixture of O₂:N₂O delivered via a small nose cone. Surgery was performed as described previously¹¹. Briefly, a small craniotomy (1–2 mm) was carried out over primary visual cortex and sealed after dye injection with 1.6% agarose in HEPES-buffered artificial cerebrospinal fluid (ACSF) and a cover slip.

Dye loading and two-photon calcium imaging *in vivo*. For bulk loading of cortical neurons, the calcium-sensitive dye Oregon Green Bapta-1 AM (OGB-1 AM; Molecular Probes) was first dissolved in 4 µl DMSO containing 20% Pluronic F-127 (Molecular Probes), and further diluted (1/11) in dye buffer (150 mM NaCl, 2.5 mM KCl and 10 mM HEPES (pH 7.4)) to yield a final concentration of 0.9 mM. Sulphorhodamine 101 (SR101, 50 µM; Molecular Probes) was added to the solution to distinguish astrocytes from neurons¹³. The dye was slowly pressure-injected into the right visual cortex at a depth of 150–200 µm with a micropipette (3–5 MΩ, 3–10 psi, 2–4 min) under visual control by two-photon imaging (×10 water immersion objective, Olympus). Activity of cortical neurons was monitored by imaging fluorescence changes with a custom-built microscope and a mode-locked Ti:sapphire laser (Mai Tai, Spectra-Physics) at 830 nm through a ×40 water immersion objective (0.8 NA, Olympus). Scanning and image acquisition were controlled by custom software written in LabVIEW (National Instruments). The average laser power delivered to the brain was <50 mW.

Imaging frames of 256 × 256 pixels were acquired at 7.6 Hz, starting at ~110 µm below cortical surface, corresponding to superficial layer 2 in mouse V1. After each recording, the focal plane and imaging position was checked and realigned with the initial image if necessary. To obtain visually evoked responses from all neurons in a cortical volume of approximately 285 × 285 × 60–120 µm³, images were recorded at 9 to 18 cortical depths with a spacing of 7 µm. At the end of each experiment, fluorescent microspheres (Lumafuor) were carefully pressure-injected into the imaged volume with a glass pipette, resulting in small fluorescent landmarks (5–20 µm diameter) along the pipette track. These landmarks were used to assist in subsequent identification of the imaged region in the sliced brain (see later), as well as fine-scale registration of *in vivo* and *in vitro* image stacks.

Visual stimulation. Visual stimuli were generated using MATLAB Psychophysics Toolbox^{33,34}, and displayed on an LCD monitor (60 Hz refresh rate) positioned 20 cm from the left eye, roughly at 45 degrees to the long axis of the animal, covering ~105 × 85 degrees of visual space. At the beginning of each experiment, the appropriate retinotopic position in visual cortex was determined using small grating stimuli at 12–24 neighbouring positions. Only cortical regions in the monocular part of primary visual cortex were included in the analysis. The monitor was repositioned such that the preferred retinotopic position of most imaged neurons was roughly in the middle of the monitor. Calcium signals were measured in response to sequences of full-field grating stimuli and natural movies. Square-wave gratings (0.035 cycles per degree, 2 cycles s⁻¹, 100% contrast) drifting in eight different directions were randomly interleaved, with the grating standing for 1.4–1.9 s before moving for 0.9–1.5 s (six repetitions per grating). Naturalistic movies consisted of 30 or 40 s sequences of moving scenes compiled from David Attenborough's *Life of Mammals* (BBC) and cage scenes from a head-mounted mouse camera, adjusted to 70% mean contrast (repeated 4–7 times).

Analysis of calcium signals. Image sequences were aligned for tangential drift and analysed with custom programs written in ImageJ (NIH), MATLAB (Mathworks) and LabVIEW. Recordings with significant brain movements, vertical drift, or both were excluded from further analysis. Outlines of neurons recorded were semi-automatically defined using software written in MATLAB (Mathworks). All pixels within each ROI were averaged to give a single time course ($\Delta F/F$), which was additionally high-pass filtered at a cut-off frequency of 0.02 Hz to remove slow fluctuations in the signal.

Spike trains were inferred from calcium signals using a fast non-negative deconvolution method that approximates the maximum a posteriori spike train for each neuron, given the fluorescence observations³⁵. Performance of the algorithm was tested by cell-attached recordings performed simultaneously with calcium imaging. There was a close correspondence between inferred and recorded spike rates (mean correlation \pm s.d. = 0.82 ± 0.06 ; nine cells from four animals).

Visual responsiveness was determined by the following procedure. For all stimulus repetitions, inferred spike trains were moving-average filtered with a time window of three frames (~0.394 s). The smoothed firing rates at corresponding points of the stimulus were then treated as groups and tested for differences by one-way ANOVA. Neurons with a *P* value less than 0.05 were considered visually

responsive. This allowed neurons that exhibited consistent elevated firing during at least one period of stimulus presentation to be detected. Among cells responsive to grating stimuli, the sum of firing rates of eight frames (~1.05 s) 0.13 s after the onset of grating drift was taken as the response to each stimulus. Responses from different trials were averaged to obtain the orientation tuning curve. This orientation tuning curve was then Fourier interpolated to 360 points, and the preferred direction was determined by the angle at which the interpolated tuning curve attained its maximum. The preferred orientation was taken as the modulus of the preferred direction to 180 degrees. OSI was calculated as $(R_{\text{best}} - R_{\text{ortho}})/(R_{\text{best}} + R_{\text{ortho}})$, where R_{best} is the interpolated response to the best direction, and R_{ortho} is the average of interpolated responses to the directions orthogonal to best responding direction. DSI was calculated as $1 - R_{\text{null}}/R_{\text{best}}$, where R_{null} is the interpolated response to the angle opposite the best responding direction. When relating connection probability to orientation selectivity or direction selectivity, neurons were defined to be orientation selective if OSI > 0.4, and direction selective if DSI > 0.3. Varying these criteria from 0.2 to 0.6 did not change the results (Supplementary Fig. 3). We used Pearson's correlation coefficient to obtain pair-wise response correlations for cell pairs, using estimated spike rates. Signal correlation was calculated as the correlation coefficient of the average responses to stimulus. Noise correlation was found by subtracting the average response from the responses to each trial, and then calculating the correlation coefficient of mean-subtracted responses.

***In vitro* whole-cell recording.** We carried out imaging experiments on a total of 16 mice that were followed by patch-clamp recordings *in vitro*. After the functional properties of individual neurons had been determined *in vivo* by two-photon calcium imaging, the mouse brain was rapidly removed and dissected in ice-cold ACSF containing 125 mM NaCl, 2.5 mM KCl, 1 mM MgCl₂, 1.25 mM NaH₂PO₄, 2 mM CaCl₂, 26 mM NaHCO₃, 25 mM dextrose; osmolality 315–325 mOsm, bubbled with 95% O₂/5% CO₂, pH 7.4. Visual cortex slices (300 µm) were cut coronally (HM 650 V Vibration Microtome, MICROM) and were incubated at 34 °C for thirty minutes before they were transferred to the recording chamber. The slice containing the imaged region was identified by the presence of OGB-1 green fluorescence and the red microsphere injection site. To reveal the relative locations of cells, a detailed morphological stack of the slice was acquired with a custom-built microscope and a mode-locked Ti:sapphire laser (Chameleon, Coherent) at 830 nm through a ×16 water immersion objective (0.8 NA, Nikon). Scanning and image acquisition were controlled by custom software written in LabVIEW. Whole-cell recordings from up to four cells were carried out in regions identified by visually comparing image stacks obtained *in vivo* and *in vitro*, using red fluorescent microspheres and the pial surface as reference. At this point, the experimenter was blind to the functional identity of the recorded neurons. Recordings were carried out in 28 °C ACSF, using Multiclamp 700B amplifiers (Axon Instruments) and data were acquired using custom software³⁶ running in Igor Pro (WaveMetrics). Recording pipettes were filled with internal solution containing 5 mM KCl, 115 mM K-gluconate, 10 mM K-HEPES, 4 mM MgATP, 0.3 mM NaGTP, 10 mM Na-phosphocreatine, 0.1% w/v biocytin, 40 µM Alexa Fluor 594; osmolality 290–295 mOsm, pH 7.2. The chloride reversal potential was approximately –85.2 mV. Junction potential was not corrected for. Cells were approached under visual guidance using laser-scanning Dodt contrast imaging. After break-through, the presence of synaptic connections was tested using five suprathreshold 5-ms-long current pulses delivered as 30-Hz trains into each cell sequentially while monitoring for postsynaptic responses in the other cells, repeated at least 30 times at 15-s intervals. Postsynaptic traces were averaged, and monosynaptic excitatory connections were deemed present when there were action-potential-locked depolarizing postsynaptic potentials associated with all five presynaptic spikes that exhibited millisecond latency³⁷. Latency was measured as the time between the peak of the action potential and 5% of the EPSP. If no spike-locked depolarizing postsynaptic potentials was present, up to 60 additional repetitions were acquired to ensure the absence of a postsynaptic response. With this approach, unitary EPSPs as small as 0.015 mV have been reported previously⁵, the smallest EPSP in the present data set was 0.035 mV. PPR was calculated as the amplitude of the second evoked EPSP over that of the first one. Input resistance was monitored throughout recordings by measuring the steady-state membrane potential change due to brief –25 pA current injections. After connectivity mapping, step currents from –50 pA to 700 pA were injected at 50 pA increments and spike threshold was measured from the inflexion point of the minimally supra-threshold trace. Spike height was the difference between spike threshold and peak. Spike half-width was measured at the mean of threshold and peak. Pyramidal neurons were identified according to morphology in Alexa 594 filled image stacks (Fig. 2a), electrophysiological properties (resting membrane potential approximately –80 mV, spike half-width >1 ms, spike height ~80 mV, regular spiking pattern typical of pyramidal neurons with current injection, see Supplementary Fig. 2c–e) and, in the presence of connections, depolarizing postsynaptic potentials (Figs 2c, 3b).

Registration of *in vivo* and *in vitro* image stacks. To accurately match up *in vivo* and *in vitro* image stacks and to locate neurons of known *in vivo* functional preference, three-dimensional image registration using custom-written MATLAB software was performed after patch-clamp experiments. The two stacks containing the same region to be matched differ in rotation and translation, as well as scales along axes. In other words, taking the centre of each stack as origin, the same points in the two stacks can be related by affine transformation. This can be written as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

Where \mathbf{x} and \mathbf{y} are column vectors of coordinates in *in vivo* and *in vitro* stack, respectively, \mathbf{A} is a 3×3 matrix representing linear transformation, and \mathbf{b} is the translation. To find the affine transformation four pairs of corresponding points, $(\mathbf{x}_i, \mathbf{y}_i)$, where $i = 1, 2, 3, 4$, were manually picked from the stacks, using landmarks such as blood vessel bifurcations, fluorescent bead injections, cortical surface, and/or brightly labelled astrocytes. The relationship between the first pair of points $(\mathbf{x}_1, \mathbf{y}_1)$ is

$$\mathbf{y}_1 = \mathbf{A}\mathbf{x}_1 + \mathbf{b}$$

$$\mathbf{b} = \mathbf{y}_1 - \mathbf{A}\mathbf{x}_1$$

Substitute this into

$$\mathbf{A}[\mathbf{x}_2 \ \mathbf{x}_3 \ \mathbf{x}_4] + [\mathbf{b} \ \mathbf{b} \ \mathbf{b}] = [\mathbf{y}_2 \ \mathbf{y}_3 \ \mathbf{y}_4]$$

and let $\mathbf{x}'_i = \mathbf{x}_i - \mathbf{x}_1$, $\mathbf{y}'_i = \mathbf{y}_i - \mathbf{y}_1$, where $i = 2, 3, 4$, gives

$$\mathbf{A}[(\mathbf{x}_2 \ \mathbf{x}_3 \ \mathbf{x}_4) - (\mathbf{x}_1 \ \mathbf{x}_1 \ \mathbf{x}_1)] = [\mathbf{y}_2 \ \mathbf{y}_3 \ \mathbf{y}_4] - [\mathbf{y}_1 \ \mathbf{y}_1 \ \mathbf{y}_1]$$

$$\mathbf{A}[\mathbf{x}'_2 \ \mathbf{x}'_3 \ \mathbf{x}'_4] = [\mathbf{y}'_2 \ \mathbf{y}'_3 \ \mathbf{y}'_4]$$

$$\mathbf{A} = [\mathbf{y}'_2 \ \mathbf{y}'_3 \ \mathbf{y}'_4][\mathbf{x}'_2 \ \mathbf{x}'_3 \ \mathbf{x}'_4]^{-1}$$

Knowing the linear transformation \mathbf{A} , \mathbf{b} can also be found.

In practice, to assist the process of identifying corresponding points, after picking three pairs of points, the image stacks were both rotated such that the planes containing the three pairs of points became parallel to the x - y plane, and the *in vitro* stack was further transformed such that the stacks became roughly registered in two dimensions on the planes but not along the z -axis (Supplementary Fig. 1a, b). To do this, let \mathbf{R}_v and \mathbf{R}_t be the matrices for rotating *in vivo* and *in vitro* stacks respectively, and let $\mathbf{u} = [u_x \ u_y \ u_z]^T$ be a unit vector, the matrix for rotating a point around \mathbf{u} by angle θ (right handedly) is given by

$$\mathbf{R} = \begin{bmatrix} 0 & -u_z & u_y \\ u_z & 0 & -u_x \\ -u_y & u_x & 0 \end{bmatrix} \sin \theta + (\mathbf{I} - \mathbf{u}\mathbf{u}^T) \cos \theta + \mathbf{u}\mathbf{u}^T$$

where \mathbf{I} is the identity matrix. To rotate the *in vivo* stack such that the plane containing the first three points picked becomes parallel to the x - y plane, the vector and the angle are

$$\mathbf{u}_v = \frac{(\mathbf{x}'_2 \times \mathbf{x}'_3) \times \mathbf{e}_z}{\|(\mathbf{x}'_2 \times \mathbf{x}'_3) \times \mathbf{e}_z\|}$$

$$\theta_v = \cos^{-1} \left(\frac{(\mathbf{x}'_2 \times \mathbf{x}'_3) \cdot \mathbf{e}_z}{\|\mathbf{x}'_2 \times \mathbf{x}'_3\|} \right)$$

where $\mathbf{e}_z = [0 \ 0 \ 1]^T$, \times denotes cross product, \cdot denotes dot product and $\| \cdot \|$ denotes norm. Substituting these into the formula for rotation matrix above, we can find \mathbf{R}_v , and similarly \mathbf{R}_t can also be found. To register the two planes in two dimensions, a further linear transformation parallel to the x - y plane can be applied to the *in vitro* stack. Let \mathbf{x}'_{iR} be the x , y components of $\mathbf{R}_v \mathbf{x}'_i$, and \mathbf{y}'_{iR} be the x , y components of $\mathbf{R}_t \mathbf{y}'_i$, where $i = 2, 3$, the matrix \mathbf{M} needed for the two-dimensional transformation is given by

$$\mathbf{M}[\mathbf{y}'_{2R} \ \mathbf{y}'_{3R}] = [\mathbf{x}'_{2R} \ \mathbf{x}'_{3R}]$$

$$\mathbf{M} = [\mathbf{x}'_{2R} \ \mathbf{x}'_{3R}][\mathbf{y}'_{2R} \ \mathbf{y}'_{3R}]^{-1}$$

Therefore, the transformation \mathbf{T}_t applied to the *in vitro* stack is

$$\mathbf{T}_t = \begin{bmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \mathbf{R}_t$$

After this step, we picked one more pair of corresponding points from a plane different from the plane that contained the initial three pairs of points (Supplementary Fig. 1b, lower panel), which is necessary for $[\mathbf{x}'_2 \ \mathbf{x}'_3 \ \mathbf{x}'_4]$ and

$[\mathbf{y}'_2 \ \mathbf{y}'_3 \ \mathbf{y}'_4]$ to be invertible, \mathbf{A} can then be found and applied to the *in vivo* stack. With the affine transformation known, we could find the correspondence between points. When rotating or transforming the image stacks, trilinear interpolation was used to assign pixel values. After registration, we inspected several planes of the transformed stack containing neurons recorded *in vitro* and made use of three-dimensional relationships of nearby cells to visually verify the matching (compare Fig. 1g, h and Supplementary Fig. 1c). Among 126 pyramidal neurons patched, matching was successful for 116 while 10 failed: 3 cells were occluded in the *in vivo* stack by a blood vessel, and 7 cells did not show convincing matching in three dimensions on visual inspection.

Analysis of connection probabilities. Connection probabilities were calculated as the number of connections detected over the number of potential connections assayed. For example, with one quadruplet there are $4 \times 3 = 12$ potential connections, and if two connections were detected the corresponding figure would be $2/12$. Probabilities of unidirectional and bidirectional connections were calculated as the number of unidirectionally and bidirectionally connected pairs over the total number of pairs, respectively. To relate connectivity to functional properties, the asymptotic Cochran–Armitage test for trend was used to test for significance³⁸. Scores of $[2, 1, 0]$ and $[2, 1, 0, 1, 2]$ were used in the test to relate connection probability or probability of finding bidirectionally connected pairs to increase in ΔOri among orientation selective pairs and direction selective pairs, respectively. Scores of $[0, 1, 2, 3, 4]$ were used to relate connection probability or probability of finding bidirectionally connected pairs to the increase in signal correlation. To relate connection probability to increase in noise correlation, scores of $[0, 1, 2]$ were used.

Criteria for inclusion in data analysis. After patching, image stacks of patched neurons filled with Alexa 594 were taken and coordinates of approximate centres of neuronal somata were manually picked in a custom-written MATLAB program. The distance between the slice surface right above the patched neuron and the soma centre was taken as the depth of neuron from the slice surface. Only neuronal pairs in which both neurons were located at $>60 \mu\text{m}$ depth and with an inter-soma distance of $<50 \mu\text{m}$ were included in the analysis relating connectivity to visual functional properties. On average we patched 7.9 neurons (range: 2–14) and assayed 13.9 potential connections (range: 2–31) per slice for neuronal pairs located deeper than $60 \mu\text{m}$ in the slice and separated by less than $50 \mu\text{m}$.

Neuronal pairs closer to the slice surface are more likely to have connections severed, and we found a significant increase in the probability of finding connections with the depth of the potential presynaptic neuron ($P = 0.043$, Cochran–Armitage test) or postsynaptic neuron ($P = 0.041$, Cochran–Armitage test) (Supplementary Fig. 6a, b). However, inclusion of cell pairs closer to the slice surface (which increases the false-negative rate of connection detection), or increasing the depth criteria (which reduces sample size) in analysis did not change the main findings (Supplementary Fig. 6d–f). Between neuronal pairs located deeper than $60 \mu\text{m}$ from the slice surface, 222 connections were assayed between pairs separated by less than $50 \mu\text{m}$. We did not find a significant difference in connection probability for neuronal pairs separated by less than $25 \mu\text{m}$ compared to those spaced farther apart ($P = 0.594$, chi-squared test; Supplementary Fig. 6c).

In 18 out of 113 pairs, high-quality recording was achieved in one cell only (for example, the other cell was very depolarized/unhealthy, or the seal resistance was less than $1 \text{ G}\Omega$). As action potentials could still be evoked in both neurons, these pairs were included as pairs in which connectivity was assayed in the direction from the unhealthy cell to the healthy cell only. Data from these pairs were included in the analysis of connection probability, but not in the analysis of probability of finding bidirectional or unidirectional pairs. Analysis of intrinsic electrophysiological properties was carried out only if series resistance was less than $30 \text{ M}\Omega$.

- Smith, S. L. & Trachtenberg, J. T. Experience-dependent binocular competition in the visual cortex begins at eye opening. *Nature Neurosci.* **10**, 370–375 (2007).
- Rocheport, N. L. et al. Sparsification of neuronal activity in the visual cortex at eye-opening. *Proc. Natl Acad. Sci. USA* **106**, 15049–15054 (2009).
- Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).
- Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
- Vogelstein, J. T. et al. Fast non-negative deconvolution for spike train inference from population calcium imaging. *J. Neurophysiol.* (2010).
- Sjöström, P. J., Turrigiano, G. G. & Nelson, S. B. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron* **32**, 1149–1164 (2001).
- Debanne, D. et al. Paired-recordings from synaptically coupled cortical and hippocampal neurons in acute and cultured brain slices. *Nature Protocols* **3**, 1559–1568 (2008).
- Agresti, A. *Categorical Data Analysis* 2nd edn (Wiley InterScience, 2002).

DISC1-dependent switch from progenitor proliferation to migration in the developing cortex

Koko Ishizuka¹, Atsushi Kamiya¹, Edwin C. Oh², Hiroaki Kanki³, Saurav Seshadri¹, Jon F. Robinson², Hannah Murdoch⁴, Allan J. Dunlop⁴, Ken-ichiro Kubo⁵, Keiko Furukori¹, Beverly Huang¹, Mariela Zeledon¹, Akiko Hayashi-Takagi¹, Hideyuki Okano³, Kazunori Nakajima⁵, Miles D. Houslay⁴, Nicholas Katsanis² & Akira Sawa^{1,6}

Regulatory mechanisms governing the sequence from progenitor cell proliferation to neuronal migration during corticogenesis are poorly understood^{1–10}. Here we report that phosphorylation of DISC1, a major susceptibility factor for several mental disorders, acts as a molecular switch from maintaining proliferation of mitotic progenitor cells to activating migration of postmitotic neurons in mice. Unphosphorylated DISC1 regulates canonical Wnt signalling via an interaction with GSK3 β , whereas specific phosphorylation at serine 710 (S710) triggers the recruitment of Bardet–Biedl syndrome (BBS) proteins to the centrosome. In support of this model, loss of BBS1 leads to defects in migration, but not proliferation, whereas DISC1 knockdown leads to deficits in both. A phospho-dead mutant can only rescue proliferation, whereas a phospho-mimic mutant rescues exclusively migration defects. These data highlight a dual role for DISC1 in corticogenesis and indicate that phosphorylation of this protein at S710 activates a key developmental switch.

In the developing cerebral cortex, progenitor cells exit the cell cycle in the ventricular and subventricular zone, whereafter postmitotic neurons move towards the cortical pial surface to form laminated cortical layers. Although proteins such as NDE1 and NDEL1 have been shown to regulate these processes, the molecular mechanisms that transition the cell state from proliferation to migration are largely unknown^{1–10}.

DISC1, a susceptibility factor for a wide range of mental illnesses, including schizophrenia, mood disorders, and autism, is expressed in both neuronal progenitor cells and postmitotic neurons in the developing cerebral cortex^{11–15}. We have reported previously that DISC1 has a role in radial neuronal migration via anchoring dynein motor-related proteins to the centrosome, including NDEL1, BBS1 and BBS4, two of the proteins mutated in Bardet–Biedl syndrome (BBS)^{13,16}. In addition, an animal model that mimics the DISC1 mutation found in a large pedigree with familial psychosis showed reduced neural proliferation during cortical midneurogenesis¹⁷. More recently, DISC1 has been shown to mediate the proliferation of neuronal progenitors in the developing cortex in a Wnt/ β -catenin-dependent fashion¹². These observations suggest that DISC1 has a dual neurodevelopmental role and raise the possibility that a switch in DISC1 function might coordinate the transition from proliferation to migration during corticogenesis.

We proposed that post-translational modification would be a strong candidate to drive the transition between the two processes. Regulated phosphorylation is an effective, rapid functional switch^{18,19}. We therefore investigated whether DISC1 is phosphorylated *in vivo* and *in vitro*. First, we treated COS7 and SH-SY5Y cells with Okadaic acid, a phosphatase inhibitor, and observed a mobility shift (slower) of exogenously expressed human DISC1, which shifted back to the original state upon treatment with lambda phosphatase (Supplementary Fig. 1a). Moreover, mouse brain extracts treated with calf alkaline phosphatase displayed a downward mobility shift of endogenous DISC1 (Supplementary Fig. 1b),

whereas Okadaic acid treatment of rat cortical primary neurons induced an upward mobility shift of endogenous DISC1. Finally, in metabolic labelling, ³²P was incorporated into DISC1, which was enhanced by Okadaic acid treatment but abolished by lambda phosphatase (Supplementary Fig. 1c). These results indicate that DISC1 is a phospho-protein in the brain.

To identify likely phosphorylated residues in DISC1, we performed mass spectrometry on exogenously expressed human DISC1 isolated from COS7 cells, treated with or without Okadaic acid. The spectrometric profile indicated three probable phosphorylation sites in DISC1 in Okadaic-acid-treated cells: threonine 50 (T50), serine 58 (S58) and S713 (Supplementary Fig. 2a). Among them, only S58 and S713 are conserved in mouse. To confirm these, we performed site-directed mutagenesis of DISC1 followed by an *in vitro* phosphorylation assay; we found that both PKA and CDK5 phosphorylated a glutathione-S-transferase (GST)-tagged C-terminal fragment of human DISC1 (amino acids 598–854). A phospho-dead mutation at S713 to alanine (A713) in human C-terminal DISC1 abolished phosphorylation, as did the orthologous S710A mutation in mouse DISC1 (Supplementary Fig. 2b). Consistent with these findings, an antibody generated against a phospho-peptide at S710 for mouse DISC1 (pS710 Ab) detected the selective immunoreactivity from extracts of HEK293 cells into which wild-type DISC1, but not phospho-dead A710-DISC1, is expressed with a catalytic subunit of PKA (also known as PRKACA; Supplementary Fig. 2c). Furthermore, pS710 Ab detects phospho-mimic mutant DISC1 (E710-DISC1: with serine replaced by glutamic acid), but cannot detect either wild-type or A710-DISC1, in the absence of active PKA (Supplementary Fig. 2c). Similar experiments showed that S58 in an N-terminal human DISC1 fragment (amino acids 1–348) was phosphorylated by PKA (Supplementary Fig. 2d).

To determine how phosphorylation of DISC1 influences signalling, we examined known interactions of DISC1, including BBS1, BBS4, NDE1 and NDEL1. We observed significantly enhanced interactions of BBS1 and BBS4 with wild-type DISC1, but not with the phospho-dead mutant A710-DISC1, upon treatment with Okadaic acid in neuronal cells (Fig. 1a and Supplementary Fig. 3a). Enhanced binding of DISC1 with BBS1 was also observed by a phospho-mimic E710-DISC1, even without the presence of Okadaic acid (Fig. 1b and Supplementary Fig. 3b). This enhancement is specific to BBS proteins, but not to NDE1 or NDEL1 (Fig. 1a and Supplementary Fig. 3c). Notably, the effect on the DISC1–BBS interaction is specific to the S710 residue; a S58A mutation did not affect DISC1–BBS1 protein interaction (Supplementary Fig. 3d).

Recruitment of BBS proteins by DISC1 to the centrosome is known to underlie neuronal migration, a key mechanism of corticogenesis¹⁶. We therefore asked whether the observed phospho-regulated DISC1–BBS1 interaction affects the centrosomal recruitment of BBS1. In cortical primary neurons transfected with E710-DISC1, we found BBS1

¹Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287 USA. ²Center for Human Disease Modeling and Departments of Cell Biology and Pediatrics, Duke University, Durham North Carolina 27710, USA. ³Departments of Physiology and ⁴Anatomy, Keio University School of Medicine, Tokyo 160-8582, Japan. ⁵Molecular Pharmacology Group, Institute of Neuroscience and Psychology, CMVLS, University of Glasgow, Glasgow G12 8QQ, UK. ⁶Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA.

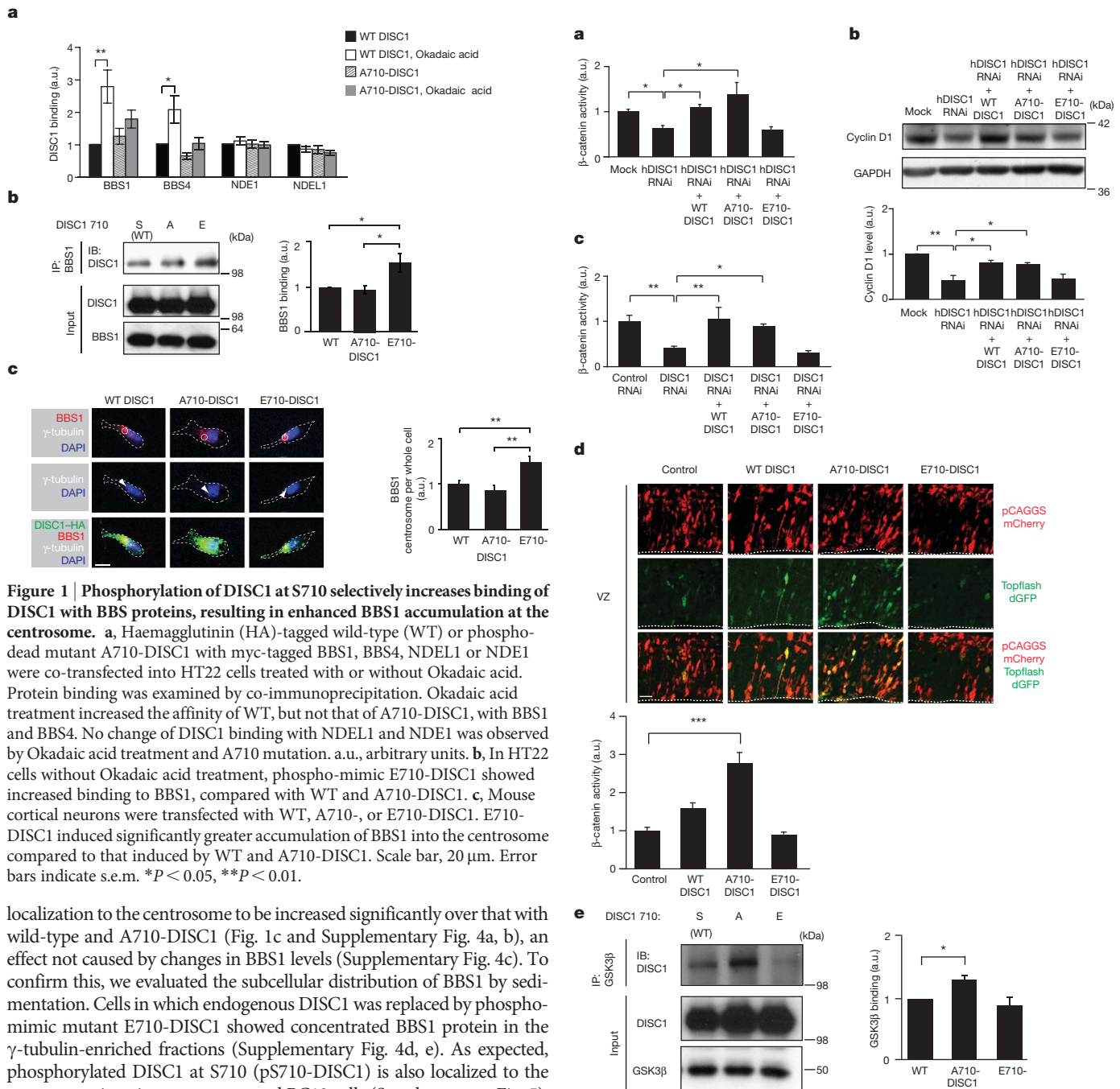


Figure 1 | Phosphorylation of DISC1 at S710 selectively increases binding of DISC1 with BBS proteins, resulting in enhanced BBS1 accumulation at the centrosome. **a**, Haemagglutinin (HA)-tagged wild-type (WT) or phospho-dead mutant A710-DISC1 with myc-tagged BBS1, BBS4, NDEL1 or NDE1 were co-transfected into HT22 cells treated with or without Okadaic acid. Protein binding was examined by co-immunoprecipitation. Okadaic acid treatment increased the affinity of WT, but not that of A710-DISC1, with BBS1 and BBS4. No change of DISC1 binding with NDEL1 and NDE1 was observed by Okadaic acid treatment and A710 mutation. a.u., arbitrary units. **b**, In HT22 cells without Okadaic acid treatment, phospho-mimic E710-DISC1 showed increased binding to BBS1, compared with WT and A710-DISC1. **c**, Mouse cortical neurons were transfected with WT, A710-, or E710-DISC1. E710-DISC1 induced significantly greater accumulation of BBS1 into the centrosome compared to that induced by WT and A710-DISC1. Scale bar, 20 μ m. Error bars indicate s.e.m. * P < 0.05, ** P < 0.01.

localization to the centrosome to be increased significantly over that with wild-type and A710-DISC1 (Fig. 1c and Supplementary Fig. 4a, b), an effect not caused by changes in BBS1 levels (Supplementary Fig. 4c). To confirm this, we evaluated the subcellular distribution of BBS1 by sedimentation. Cells in which endogenous DISC1 was replaced by phospho-mimic mutant E710-DISC1 showed concentrated BBS1 protein in the γ -tubulin-enriched fractions (Supplementary Fig. 4d, e). As expected, phosphorylated DISC1 at S710 (pS710-DISC1) is also localized to the centrosome in primary neurons and PC12 cells (Supplementary Fig. 5).

The canonical Wnt pathway is a key regulator of progenitor cell proliferation in the developing cortex²⁰. Moreover, several studies have shown that the centrosome/basal body in postmitotic cells acts as a negative regulator of canonical Wnt signalling, because suppression of BBS1 and BBS4 leads to the aberrant activation of β -catenin signalling^{21,22}. We therefore proposed that the phosphorylation-enhanced DISC1–BBS1 binding might titrate DISC1 away from a Wnt/ β -catenin activity and thus contribute to the switch from neuronal progenitor proliferation to neuronal migration. To test this, we examined whether phosphorylation of DISC1 at S710 influenced canonical Wnt signalling by using the established Wnt reporter cell line, HEK 293T Super 8x TOPFlash^{23,24}. Knockdown of DISC1 suppressed Wnt/ β -catenin transcriptional activity upon stimulation by WNT3A (Fig. 2a). Importantly, the rescue of these phenotypes was dependent on the phosphorylation status of DISC1: co-expression of the phospho-dead mutant A710-DISC1 resulted in efficient rescue, whereas the phospho-mimic mutant E710-DISC1 failed completely (Fig. 2a). We also evaluated the expression

Figure 2 | Non-phosphorylated DISC1 at S710 activates β -catenin signalling via its interaction with GSK3 β . **a**, *In vitro* luciferase assay showed that DISC1 knockdown suppressed β -catenin-dependent activity, which was rescued by WT and phospho-dead A710-DISC1, but not by phospho-mimic E710-DISC1. hDISC1, human DISC1. **b**, Expression of Cyclin D1 was downregulated by DISC1 RNAi, which was normalized by WT and A710-DISC1, but not by E710-DISC1. **c**, Super 8x TOPFlash and pRL SV40 plasmids together with various constructs were injected *in utero* at E13 brain and were analysed at E15. Knockdown of DISC1 suppressed β -catenin-dependent activity, which was rescued by WT and A710-DISC1, but not by E710-DISC1. **d**, *In utero* assay to monitor activation of the β -catenin pathway with the TOPdGFP-CAG mCherry assay system. Brains were analysed at E14. Relative β -catenin activity was assessed by the ratio of the number of GFP-positive cells to the number of mCherry-positive cells; significant upregulation of β -catenin signalling activity was evident in the brain injected with A710-DISC1. VZ, ventricular zone. Scale bar, 20 μ m. **e**, Increased binding of GSK3 β with A710-DISC1 compared to WT DISC1 in HT22 cells by co-immunoprecipitation (IP) and immunoblot (IB). Error bars indicate s.e.m. * P < 0.05, ** P < 0.01, *** P < 0.001.

of cyclin D1, a known β -catenin transcriptional target. Our data were consistent with the reporter assays: cyclin D1 levels, which were suppressed by DISC1 RNA interference (RNAi), were rescued by wild-type and A710-DISC1, but not by E710-DISC1 (Fig. 2b).

We next assessed the impact of DISC1 RNAi on Wnt signalling *in vivo* by *in utero* gene transfer. Consistent with our *in vitro* data, knock-down of DISC1 by RNAi co-injected with Super 8x TOPFlash and pRL SV40 expression constructs at embryonic day 13 (E13) induced a significant reduction of Wnt/ β -catenin signalling activity at E15, which was rescued by wild-type and A710-DISC1, but not by E710-DISC1 (Fig. 2c). When we compared the effects of overexpression of wild-type and mutant DISC1 on β -catenin transcriptional activity by using a second reporter construct which expresses a destabilized GFP variant under the control of a β -catenin-responsive promoter and constitutive CAG-promoter-driven mCherry in tandem (TOPdGFP-CAG mCherry)^{25,26}, we also observed significant upregulation of active β -catenin, as indicated by enhanced green fluorescent protein (GFP) expression with A710-DISC1, but not with E710-DISC1 (Fig. 2d). In parallel, protein binding of GSK3 β with DISC1 was augmented by expression of A710-DISC1, but not by E710-DISC1 (Fig. 2e), which was also supported by an *in vitro* binding assay (Supplementary Fig. 6).

To explore the physiological relevance of these data during cortical development, we assessed the levels of pS710-DISC1 in the developing cortex at E14 and E18 with our pS710 antibody (Supplementary Figs 2c and 5a). As expected, the levels of pS710-DISC1 were greater at E18 (when neuronal migration is prominent) compared with those at E14 (when progenitor proliferation is prominent) (Fig. 3a), whereas immunohistochemistry with pS710 antibody indicated selective staining in the cortical plate/intermediate zone, but not in the ventricular/subventricular zone (Fig. 3b and Supplementary Fig. 7). We then

quantified the relative protein binding of DISC1 with interactors in mouse brain lysates at E14 and E18 by co-immunoprecipitation. We found that the affinity between DISC1 and BBS1 increased from E14 to E18, whereas the DISC1–GSK3 β affinity decreased proportionally during this period (Fig. 3c). Moreover, we observed negligible binding of GSK3 β and pS710-DISC1, whereas protein binding of BBS1 and pS710-DISC1 augmented significantly at E18 compared to E14 (Fig. 3c). We propose the following model: during mid-embryonic stages when progenitor cell proliferation is prominent (including E14), unphosphorylated DISC1 at S710 binds with GSK3 β more tightly and regulates cell proliferation; in contrast, during later embryonic stages when cell cycle exit of progenitors and the following neuronal migration become predominant, pS710-DISC1 dissociates from GSK3 β and switches its role to the recruitment of BBS1 to the centrosome, activating neuronal migration. To test this model directly, we flow-sorted homogeneous populations of mitotic progenitor cells and post-mitotic neuronal cells from the developing cortex of transgenic mice expressing a nestin-promoter-driven Kusabira-Orange and doublecortin (DCX)-promoter-driven enhanced green fluorescent protein (EGFP)²⁷. We observed an increased abundance of pS710-DISC1 and a reduced affinity for GSK3 β with a concomitant increase in the binding for BBS1 both for total DISC1 and pS710-DISC1 in the EGFP-positive post-mitotic cells compared with Kusabira-Orange-positive progenitor cells (Fig. 3d and Supplementary Fig. 8).

Another prediction would be that loss of BBS1 should lead to aberrant migration, but not proliferation. In *Bbs1* knockout mice²⁸, 5-bromodeoxyuridine (BrdU) labelling and scoring at E15 blind to genotype showed that proliferation and cell cycle exit were indistinguishable between null and wild-type littermates (Fig. 3e). By contrast, both CUX1- (marking layers II–IV)- and CTIP2 (also known as

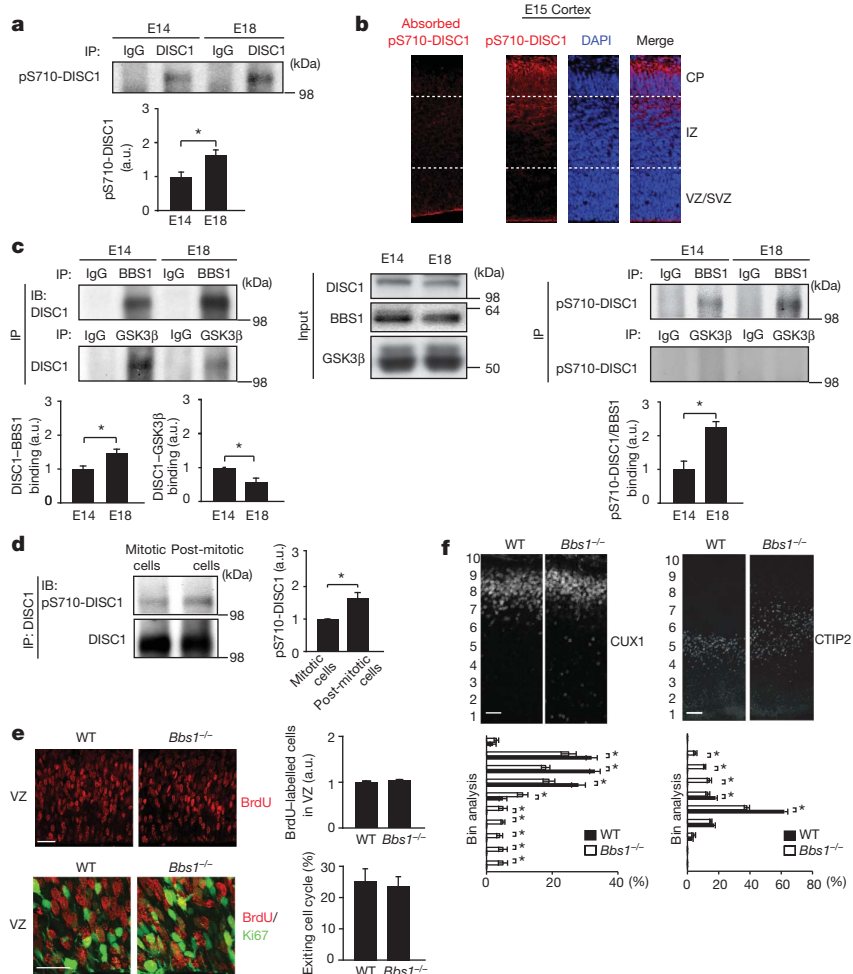


Figure 3 | Binding affinity of DISC1–GSK3 β to DISC1–BBS1 depends on developmental stage: roles of BBS1 in corticogenesis. **a**, Levels of phosphorylated DISC1 at S710 (pS710-DISC1) at E14 and E18. Immunoprecipitation with a pan-DISC1 antibody was conducted, and immunoprecipitates were analysed with pS710 phospho-specific antibody or another pan-DISC1 antibody. Total levels of DISC1 were unchanged (see Fig. 3c), whereas levels of pS710-DISC1 were significantly higher at E18 compared to E14. **b**, pS710-DISC1 was prominent only in the cortical plate (CP)/intermediate zone (IZ), but not in ventricular/subventricular zone (VZ/SVZ) at E15, suggesting that phosphorylation may occur preferentially in post-mitotic neurons, but not in progenitors. Scale bar, 20 μ m. **c**, Binding of DISC1–BBS1 and DISC1–GSK3 β was assessed in E14 and E18 mouse brains by co-immunoprecipitation. DISC1–BBS1 binding was increased during this period, whereas DISC1–GSK3 β binding was decreased. Furthermore, pS710-DISC1–BBS1 binding was significantly greater at E18 compared with E14; minimal binding of pS710-DISC1 with GSK3 β was observed. **d**, Levels of pS710-DISC1 in mitotic progenitors and post-mitotic neurons were assessed by immunoprecipitation. Total levels of DISC1 were unchanged, whereas levels of pS710-DISC1 were significantly higher in post-mitotic neurons compared to mitotic progenitors. **e**, No appreciable differences in the progenitor cell proliferation of the cortex between *Bbs1*^{−/−} mice and wild-type littermates, as determined by BrdU incorporation and cell cycle exit index. Scale bar, 20 μ m. **f**, Aberrant radial migration in *Bbs1*^{−/−} mice compared to WT littermates, assayed by bin analysis with the brains at postnatal day 0 (P0). CUX1 (layers II–IV) and CTIP2 (layer V) were used as indicators. Scale bar, 50 μ m. Error bars indicate s.e.m. **P* < 0.05.

BCL11B, marking layer V)-positive cells showed aberrant positioning in *Bbs1*^{-/-} mice (Fig. 3f); a migration defect in superficial layers was also observed, when BrdU or GFP was injected at E15 and the final positioning of the late-born superficial layer neurons was analysed at postnatal day 0 (P0) and E19 (Supplementary Fig. 9).

To obtain further *in vivo* evidence, we examined how the phosphorylation status at S710 of DISC1 differentially regulates proliferation and neuronal migration in mid- (E13–15) and later- (E15–19) embryonic stages, respectively, by *in utero* gene transfer (Fig. 4 and Supplementary Figs 10–16). Consistent with previous reports¹², DISC1 knockdown at E13 leads to several deficits and altered cell fate associated with the progenitor proliferation (Fig. 4a and Supplementary Figs 11 and 12a–c). Of note, DISC1 knockdown did not induce marked differences in N-cadherin staining (Supplementary Fig. 12d). In contrast, several groups have reported that suppression of DISC1 at a later time point, such as E15, led to delayed neuronal migration^{13,16}, a phenotype which, if our model is correct, should be dependent on pS710-DISC1. Consistently, we efficiently rescued the DISC1 RNAi-induced migration phenotype either with wild-type DISC1, or with the phosphomimic mutant E710-DISC1, but not with the phospho-dead mutant A710-DISC1 (Fig. 4b and Supplementary Fig. 13), which is in sharp contrast to the effects of E710-DISC1 and A710-DISC1 on progenitor cell proliferation at earlier time points (Fig. 4a and Supplementary Fig. 11). It is unlikely that the migratory defects are caused by disturbed radial scaffold formation, as we observed no significant effect of DISC1 suppression in radial fibre elongation (Supplementary Fig. 14). Finally, expression of a dominant-negative CDK5 (ref. 29) could induce migration defects that phenocopy the DISC1 RNAi phenotypes and

are ameliorated by co-expression of E710-DISC1, but not by A710-DISC1 or by wild-type DISC1 (Fig. 4c).

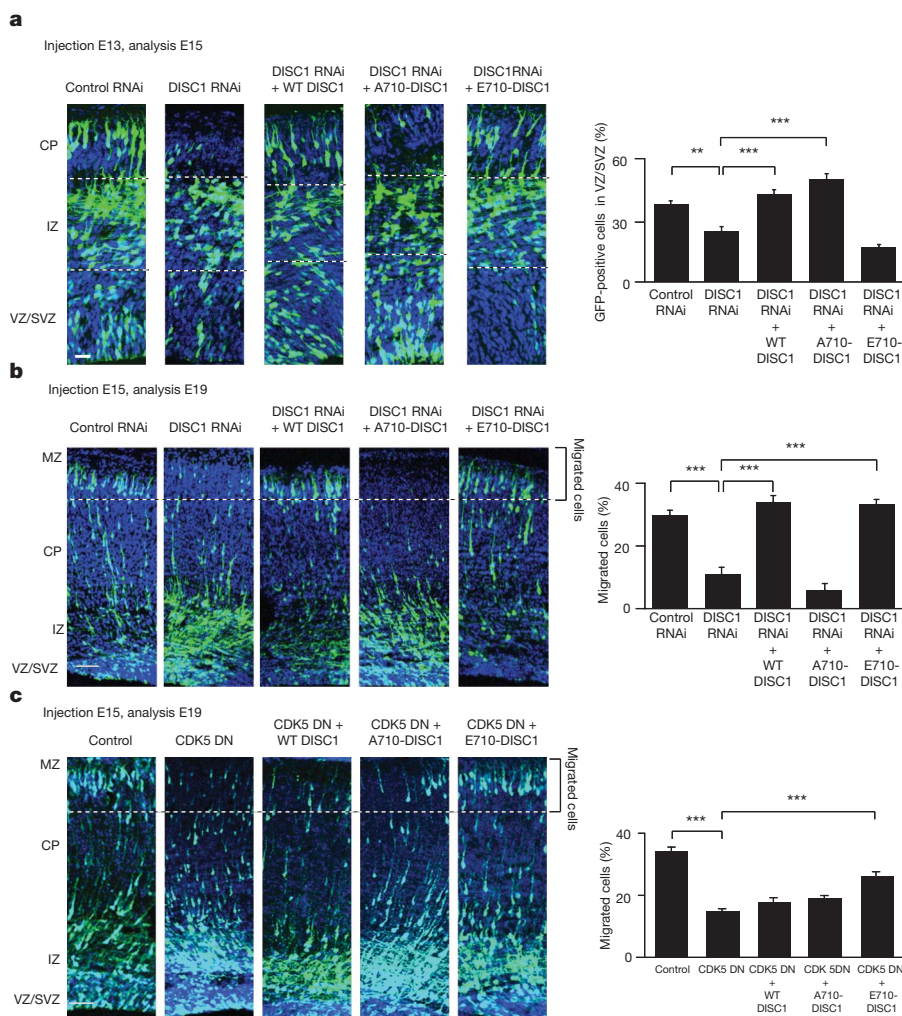
The main finding of our study is that DISC1 is a dynamic protein that acts as a molecular switch between two key stages of cortical development, cell proliferation and neuronal migration. Consistent with our model, *Bbs1*^{-/-} mice display intact proliferation but abnormal migration that was not rescued by wild-type DISC1 (Supplementary Fig. 15). Also, the migration deficit in *Bbs1* knockout mice is more modest compared to that of DISC1 suppression *in utero*, suggesting the presence of other mediators downstream of DISC1.

We cannot exclude the possibility that abnormal positioning in radial migration might be influenced by the observed proliferation deficit. However, DCX-promoter-driven wild-type and E710-DISC1 expressed only in the post-mitotic period can also rescue the migration deficit, indicating that the abnormal neuronal positioning caused by DISC1 RNAi is not a result of impaired proliferation (Supplementary Fig. 16). Two further lines of evidence support this: first, when we injected DISC1 RNAi into E15 brains and analysed them at E19, cell proliferation had mostly ended⁴; second, depletion of BBS1 is sufficient for abnormal neuronal positioning *in vivo* (Fig. 3f and Supplementary Fig. 9).

Our data raise a number of new questions. For instance, it is unclear why the BBS proteins have a more selective role for the switch to neuronal migration. Recent data have suggested that several BBS proteins may be crucial for the interpretation of planar cell polarity pathway signalling, while antagonizing canonical Wnt/ β -catenin signalling²¹. Moreover, biochemical kinetics of DISC1–GSK3 β –BBS proteins in association with this allosteric regulation by S710 phosphorylation may also be important.

Figure 4 | Suppression of DISC1 leads to phospho-dependent defects in cell proliferation and neuronal migration: implication of CDK5.

a, Various constructs were injected *in utero* at E13 and analysed at E15. In brains with DISC1 RNAi, the percentage of GFP-positive cells in the ventricular and subventricular zones (VZ/SVZ) was lower compared to brains with control RNAi, which were rescued by WT DISC1 and phospho-dead A710-DISC1 but not by phospho-mimic E710-DISC1. Scale bar, 20 μ m. **b**, *In utero* gene transfer at E15, analysed at E19. The percentage of GFP-positive cells reaching the superficial layers of the cortex (bins 9 and 10; migrated cells) was examined. Silencing of DISC1 induced delayed neuronal migration, rescued by WT and E710-DISC1 but not by A710-DISC1. Scale bar, 50 μ m. **c**, WT, A710-, or E710-DISC1 constructs, together with a dominant-negative CDK5 construct (CDK5 DN) were electroporated at E15 and their effects on migration were analysed at E19. Migration defects elicited by CDK5 DN were rescued significantly by co-expression with E710-DISC1, but not with A710-DISC1 or with WT DISC1. Scale bar, 50 μ m. Error bars indicate s.e.m. ** P < 0.01, *** P < 0.001.



Finally, it is crucial to consider how disturbances in the DISC1-dependent switch mechanism might have a clinical impact. We speculate that disturbance of this switch mechanism may contribute to hypertrophic and disturbed corticogenesis observed in brains of patients with autism. Although schizophrenia has an onset in young adulthood, the initial pathological insults occur during neurodevelopment¹⁴. It is possible that disturbances of this molecular switch might also underlie this pathology.

METHODS SUMMARY

Mice. C57/BL6 pregnant female mice were purchased from Charles River for *in utero* gene transfer. *Bbs1* knockout mice were backcrossed four generations into the C57/BL6 background and characterized elsewhere²⁸.

In utero electroporation and immunohistochemistry. *In utero* electroporation was performed as described (see refs 12, 13, 16, 30 and Methods). The RNAi plasmids that had been fully characterized in publications from other groups and ours (see refs 12, 13, 16 and Methods), were electroporated at E13 or E15. Rescue experiments were conducted by a combination of RNAi plasmid (0.1 $\mu\text{g } \mu\text{l}^{-1}$ in 1 μl) with wild-type or mutant DISC1 expression plasmid (2.5 $\mu\text{g } \mu\text{l}^{-1}$ in 1 μl). Coronal slices of developing cerebral cortex were prepared as described^{13,30}. Briefly, the brains were fixed with 4% paraformaldehyde and sectioned with a cryostat at 20 μm on E14, E15 and E19. Nuclei were labelled with DAPI (Invitrogen). Slice images were acquired with a confocal microscope (Zeiss LSM510 Meta).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 July 2010; accepted 21 January 2011.

Published online 6 April 2011.

1. Ayala, R., Shu, T. & Tsai, L. H. Trekking across the brain: the journey of neuronal migration. *Cell* **128**, 29–43 (2007).
2. Bielas, S., Higginbotham, H., Koizumi, H., Tanaka, T. & Gleeson, J. G. Cortical neuronal migration mutants suggest separate but intersecting pathways. *Annu. Rev. Cell Dev. Biol.* **20**, 593–618 (2004).
3. Gupta, A., Tsai, L. H. & Wynshaw-Boris, A. Life is a journey: a genetic look at neocortical development. *Nature Rev. Genet.* **3**, 342–355 (2002).
4. Dehay, C. & Kennedy, H. Cell-cycle control and cortical development. *Nature Rev. Neurosci.* **8**, 438–450 (2007).
5. Götz, M. & Huttner, W. B. The cell biology of neurogenesis. *Nature Rev. Mol. Cell Biol.* **6**, 777–788 (2005).
6. Kriegstein, A., Noctor, S. & Martinez-Cerdeno, V. Patterns of neural stem and progenitor cell division may underlie evolutionary cortical expansion. *Nature Rev. Neurosci.* **7**, 883–890 (2006).
7. Mochida, G. H. & Walsh, C. A. Genetic basis of developmental malformations of the cerebral cortex. *Arch. Neurol.* **61**, 637–640 (2004).
8. Leone, D. P., Srinivasan, K., Chen, B., Alcamo, E. & McConnell, S. K. The determination of projection neuron identity in the developing cerebral cortex. *Curr. Opin. Neurobiol.* **18**, 28–35 (2008).
9. Shu, T. *et al.* Ndel1 operates in a common pathway with LIS1 and cytoplasmic dynein to regulate cortical neuronal positioning. *Neuron* **44**, 263–277 (2004).
10. Reiner, O. *et al.* Isolation of a Miller–Dieker lissencephaly gene containing G protein β -subunit-like repeats. *Nature* **364**, 717–721 (1993).
11. Schurov, I. L., Handford, E. J., Brandon, N. J. & Whiting, P. J. Expression of disrupted in schizophrenia 1 (DISC1) protein in the adult and developing mouse brain indicates its role in neurodevelopment. *Mol. Psychiatry* **9**, 1100–1110 (2004).
12. Mao, Y. *et al.* Disrupted in schizophrenia 1 regulates neuronal progenitor proliferation via modulation of GSK3 β / β -catenin signaling. *Cell* **136**, 1017–1031 (2009).
13. Kamiya, A. *et al.* A schizophrenia-associated mutation of DISC1 perturbs cerebral cortex development. *Nature Cell Biol.* **7**, 1167–1178 (2005).
14. Jaaro-Peled, H. *et al.* Neurodevelopmental mechanisms of schizophrenia: understanding disturbed postnatal brain maturation through neuregulin-1–ErbB4 and DISC1. *Trends Neurosci.* **32**, 485–495 (2009).
15. Chubb, J. E., Bradshaw, N. J., Soares, D. C., Porteous, D. J. & Millar, J. K. The DISC locus in psychiatric illness. *Mol. Psychiatry* **13**, 36–64 (2008).
16. Kamiya, A. *et al.* Recruitment of PCM1 to the centrosome by the cooperative action of DISC1 and BBS4: a candidate for psychiatric illnesses. *Arch. Gen. Psychiatry* **65**, 996–1006 (2008).
17. Shen, S. *et al.* Schizophrenia-related neural and behavioral phenotypes in transgenic mice expressing truncated Disc1. *J. Neurosci.* **28**, 10893–10904 (2008).
18. Greengard, P. *et al.* The DARPP-32/protein phosphatase-1 cascade: a model for signal integration. *Brain Res. Brain Res. Rev.* **26**, 274–284 (1998).
19. Xie, Z., Sanada, K., Samuels, B. A., Shih, H. & Tsai, L. H. Serine 732 phosphorylation of FAK by Cdk5 is important for microtubule organization, nuclear movement, and neuronal migration. *Cell* **114**, 469–482 (2003).
20. Chenn, A. & Walsh, C. A. Regulation of cerebral cortical size by control of cell cycle exit in neural precursors. *Science* **297**, 365–369 (2002).
21. Gerdes, J. M. *et al.* Disruption of the basal body compromises proteasomal function and perturbs intracellular Wnt response. *Nature Genet.* **39**, 1350–1360 (2007).
22. Corbit, K. C. *et al.* Kif3a constrains β -catenin-dependent Wnt signalling through dual ciliary and non-ciliary mechanisms. *Nature Cell Biol.* **10**, 70–76 (2008).
23. Veeman, M. T., Slusarski, D. C., Kaykas, A., Louie, S. H. & Moon, R. T. Zebrafish prickles, a modulator of noncanonical Wnt/Fz signaling, regulates gastrulation movements. *Curr. Biol.* **13**, 680–685 (2003).
24. Korinek, V. *et al.* Constitutive transcriptional activation by a β -catenin–Tcf complex in APC^{−/−} colon carcinoma. *Science* **275**, 1784–1787 (1997).
25. Yokota, Y. *et al.* The adenomatous polyposis coli protein is an essential regulator of radial glial polarity and construction of the cerebral cortex. *Neuron* **61**, 42–56 (2009).
26. Zhang, J. *et al.* Cortical neural precursors inhibit their own differentiation via N-cadherin maintenance of β -catenin signaling. *Dev. Cell* **18**, 472–479 (2010).
27. Kanki, H., Shimabukuro, M. K., Miyawaki, A. & Okano, H. “Color Timer” mice: visualization of neuronal differentiation with fluorescent proteins. *Mol. Brain* **3**, 5 (2010).
28. Kulaga, H. M. *et al.* Loss of BBS proteins causes anosmia in humans and defects in olfactory cilia structure and function in the mouse. *Nature Genet.* **36**, 994–998 (2004).
29. Ohshima, T. *et al.* Cdk5 is required for multipolar-to-bipolar transition during radial neuronal migration and proper dendrite development of pyramidal neurons in the cerebral cortex. *Development* **134**, 2273–2282 (2007).
30. Tabata, H. & Nakajima, K. Efficient *in utero* gene transfer system to the developing mouse brain using electroporation: visualization of neuronal migration in the developing cortex. *Neuroscience* **103**, 865–872 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Y. Lema for preparing figures and organizing the manuscript and P. Talalay for critical reading of this manuscript. We appreciate J. M. Gerdes, A. K. Mustafa, N. Shahani, T. Boronina, D. Chen and R. N. Cole for scientific discussions and technical support. We thank T. Tomoda, N. Sugiyama, M. Hasegawa, Q. Lu, E. S. Anton and A. Chenn for providing us with constructs. This work was supported by USPHS grants of MH-084018 Silvio O. Conte center (A.S.), MH-069853 (A.S.), MH-085226 (A.S.), MH-088753 (A.S.), MH-091230 (A.K.), HD-04260 (N.K.), DK-072301 (N.K.), and DK-075972 (N.K.); grants from Stanley and RUSK foundations and from Maryland Stem Cell Research Fund (A.S.); grants from NARSAD and S-R foundations (A.S. and A.K.); grants from the Macular Vision Research Foundation and the Foundation for Fighting Blindness as well as the Distinguished George W. Brumley Professorship (N.K.); a grant from Health Labor Sciences (K.-i.K.); grants from Strategic Research Program for Brain Sciences (K.N.), MEXT (K.N.), Takeda (K.N.) and PMAC-PSJ (K.N.); Fight for Sight Postdoctoral Fellowship (E.O.); grant from the Medical Research Council, UK (G0600765; M.D.H.).

Author Contributions K.I. and A.S. conceived the general ideas for this study. K.I., A.K., E.C.O., N.K. and A.S. designed experiments. K.I., A.K., E.C.O. and J.F.R. performed the experiments and data analysis with assistance from H.K., S.S., H.M., A.J.D., K.-i.K., K.F., B.H., M.Z., A.H.-T., H.O., K.N. and M.D.H.; K.I., A.K., E.C.O., M.D.H., N.K. and A.S. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.S. (asawa1@jhmi.edu) or N.K. (katsanis@cellbio.duke.edu).

METHODS

Mice. C57/BL6 pregnant female mice were purchased from Charles River for *in utero* gene transfer. *Bbs1* knockout mice were backcrossed four generations into the C57/BL6 background and characterized elsewhere²⁸.

Plasmids. We used the RNAi constructs to DISC1 that have been established in the previous publications from other groups and ours^{12,13,16,31–33}. A scrambled sequence without homology to any known mRNA was used as control RNAi. The rescue constructs contained three nucleotide alterations in the RNAi target sequence to avoid silencing by RNAi. DCX promoter-driven DISC1 expression constructs (DCX-DISC1) were made by replacing the EGFP-expressing cDNA sequence in DCX-EGFP³⁴ with WT or mutant DISC1-expressing cDNA sequences.

Antibodies. A polyclonal human DISC1 antibody (hExon2 Ab) was raised against amino acids 285–298 in human DISC1. hExon2 Ab detects signals of exogenously expressed human DISC1 in COS7 cell and endogenous DISC1 in SH-SY5Y cell lysates, which were abolished by pre-absorption with original antigen. A polyclonal antibody against the phosphopeptide C-LQLQEAGSpSPHADE (amino acids 702–716 of mouse DISC1) (pS710 Ab) was generated. These signals were abolished by pre-absorption with the original phosphopeptide. An antibody against BBS1 (ref. 35) and antibodies against DISC1 (refs 11, 36) have been described. The following commercial antibodies were also used: mouse monoclonal antibodies against γ -tubulin (Sigma), HA-tag (Covance), GST-tag (Covance), myc-tag (Santa Cruz), GSK3 α/β (Santa Cruz), GAPDH (Santa Cruz), Pax6 (Iowa Hybridoma Bank), cyclin D1 (Cell Signaling), N-cadherin (Invitrogen), BrdU (Chemicon), and RC2 antibody (Iowa Hybridoma Bank); mouse polyclonal antibody against BBS1 (Novus); rat monoclonal antibody against HA-tag (Roche); goat polyclonal antibodies against ER81 (Santa Cruz), and Doublecortin (Santa Cruz); rabbit polyclonal antibodies against TBR2 (Abcam) and γ -tubulin (Sigma); and rabbit monoclonal antibody against Ki67 (NeoMarkers).

Cells and transfection. COS7, SH-SY5Y, and HT22 cells were grown in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum. PC12 cells were maintained in DMEM containing 10% fetal bovine serum and 5% horse serum. FuGENE 6 (Roche Applied Sciences) was used for transfection to COS7 cells. SH-SY5Y, HT22 and PC12 cells were transfected with Lipofectamine 2000 (Invitrogen). When used, the phosphatase inhibitor Okadaic acid (0.5 μ M; Calbiochem) was added 2 h before cells were collected. Dissociated cortical neuron cultures were prepared as described previously³³.

Immunoprecipitation. Cells or tissue were lysed in lysis buffer (150 mM NaCl, 50 mM Tris-HCl, pH 7.5, 1% Triton X-100) containing protease inhibitor mixture (Roche Applied Sciences) and phosphatase inhibitor cocktail (Sigma). Lysates were sonicated, cell debris was cleared by centrifugation, and the soluble fraction was subjected to immunoprecipitation as described previously¹³.

Immunofluorescent staining. When transfected, 1 to 2 days after transfection, primary cortical neurons, PC12, or HT22 cells were fixed with ice-cold methanol at -20°C for 15 min. After blocking with 0.1% Triton X-100 and 2% normal goat serum in PBS, cells were incubated with primary antibodies at 4°C overnight, followed by reaction with secondary antibodies conjugated to Rhodamine Red-X, Cy5 (Jackson Immuno Research), and Alexa488 (Molecular Probe) for 1 h. DAPI (Invitrogen) was used to visualize nuclei. Confocal microscopy (Zeiss LSM 510 Meta) was used for epifluorescent image collection. The distribution of BBS1 at the centrosome in cells was quantified as described¹⁶. Briefly, a circle with 3- μ m diameter was drawn centring on the γ -tubulin and defined as the area including the centrosome. Whole-cell area was determined by distribution of overexpressed DISC1-HA with HA staining. The intensity of BBS1 staining in the whole cell area versus that in the centrosome area was quantified with Metamorph (Molecular Devices) for all experimental groups. The intensity ratio of the signal of more than 30 cells per group was analysed in three independent experiments in a blinded manner.

***In vitro* β -catenin activity assays with luciferase reporter system.** Luciferase reporter system assays were carried out as described²¹. HEK293T cells stably expressing pTOPFlash reporter were seeded in 24-well plates at a density of 10^4 cells per well. After 18–24 h, reporter plasmid and/or *Renilla* luciferase cDNA in an SV40 vector and the plasmid of interest were transfected in six wells by using the FuGENE 6 (Roche) optimized transfection protocol. pRL SV40 (*Renilla* luciferase) was used as an internal control. When applicable, after 24 h we treated three wells from each plate with Wnt3a-enriched medium that had been aspirated from Wnt3a/L cells and sterile filtered before being applied to the luciferase assay. Cells were lysed and luciferase activity was measured 48 h after start of stimulation by using the Promega Dual Luciferase Reporter Assay System (E1910) and a FLUOstar Luminometer (BMG Technologies). Each assay was repeated at least twice to ensure reproducibility of the results.

Purification of mitotic progenitors and post-mitotic neurons. Progenitor cells and post-mitotic neurons were purified by FACS from the brains of transgenic

mice expressing nestin promoter-driven Kusabira-Orange and DCX promoter-driven EGFP, respectively²⁷.

***In utero* electroporation and immunohistochemistry.** *In utero* electroporation was performed as described^{12,13,16,30–32}. The RNAi plasmids that had been fully characterized in publications from other groups and ours^{12,13,16,31–33}, were electroporated at E13 or E15. Rescue experiments were conducted by a combination of RNAi plasmid (0.1 $\mu\text{g } \mu\text{l}^{-1}$ in 1 μl) with wild-type or mutant DISC1 expression plasmid (2.5 $\mu\text{g } \mu\text{l}^{-1}$ in 1 μl). Coronal slices of developing cerebral cortex were prepared as described^{13,30}. Briefly, the brains were fixed with 4% paraformaldehyde and sectioned with a cryostat at 20 μm on E14, E15 and E19, respectively. Nuclei were labelled with DAPI (Invitrogen). Slice images were acquired with a confocal microscope (Zeiss LSM510 Meta).

***In vivo* β -catenin activity assays with luciferase reporter system.** Super 8x TOPFlash and pRL SV40 together with expression constructs and/or RNAi constructs were electroporated at E13 and luciferase activity was measured at E15.

***In vivo* β -catenin activity assays with TOPdGFP-CAG mCherry.** A dual reporter construct expressing a destabilized GFP variant under the control of a β -catenin responsive promoter and constitutive CAG promoter-driven mCherry in tandem (TOPdGFP-CAG mCherry)^{25,26} together with expression constructs and/or RNAi constructs were electroporated at E13. After 24 h, the brains were extracted and assessed for mCherry and GFP expression.

BrdU incorporation assay. BrdU (50 mg/kg) was injected i.p. into pregnant mice 48 h after electroporation. After 2 h, brains were processed and sections were stained with anti-BrdU antibodies.

Cell cycle exit assay. BrdU (50 mg kg $^{-1}$) was injected intraperitoneally into pregnant mice 24 h after electroporation. Twenty-four hours later, brains were processed and sections were stained with anti-BrdU and anti-Ki67 antibodies.

Definition of the ventricular zone (VZ), subventricular zone (SVZ), and intermediate zone (IZ) for assessment of progenitor proliferation at E15. The VZ/SVZ boundary was defined by the segregation of PAX6- and TBR2 (also known as EOMES)-positive cells. IZ was determined as TBR2-negative and DCX-positive area. In addition, morphological characteristics were used as indicators of VZ/SVZ and SVZ/IZ boundaries in this study: VZ and SVZ were separated by existence of multipolar cells (arrowhead), and SVZ and IZ were divided by cell density detected by DAPI staining³⁷.

Quantitative bin analysis for assessment of migration at E19 or P0. To quantify the pattern of migration in E19 or P0 brains, the numbers of GFP-positive cells in the developmental cerebral cortex were counted from three independent sections. We quantified the RNAi effect on neuronal migration status by bin analysis, in which the developing cerebral cortex was divided into 10 equal spaces (10 bins) and the percentage of GFP-positive cells in each bin was determined^{13,16}. The numbers of neurons in each category from more than three independent experiments were counted in a blinded manner. Migration distance was defined as the relative distance of each cell migration (from the surface of the ventricle) to the radial thickness of the cerebral cortex where the cells were located¹⁶. The cells reaching the superficial layers of the cortex (bins 9 and 10) were examined as migrated cells.

Phosphatase treatment. Soluble proteins obtained from the cells overexpressing myc-tagged wild-type human DISC1 were incubated with an antibody against myc-tag (Santa Cruz) and Protein G Plus/Protein A agarose beads (Calbiochem) at 4°C overnight. Beads were washed in 20 mM Tris-HCl, pH 7.6 three times and in lambda phosphatase buffer (New England Biolabs) once, and phosphatase reactions were performed directly on the beads at 30°C for 2 h with lambda phosphatase (New England Biolabs) as per manufacturer's protocol. Immune complexes were then washed three times in lysis buffer, separated on SDS-PAGE, and analysed by western blotting. Mouse brains were homogenized in lysis buffer. Fifty micrograms of soluble proteins from each sample in 20 μl of dephosphorylation buffer (Roche Applied Sciences) were incubated with or without 2 μl of calf alkaline phosphatase (Roche Applied Sciences) at 37°C for 1 h. Reactions were stopped with SDS sample buffer and the samples were subjected to western blotting.

Metabolic labelling. COS7 cells expressing myc-tagged wild-type human DISC1 were metabolically labelled for 4 h at 37°C with 0.5 mCi ml $^{-1}$ [^{32}P]orthophosphate (PerkinElmer Life Sciences) in phosphate- and serum-free media with or without okadaic acid (0.5 μM ; Calbiochem). Immunoprecipitation and lambda phosphatase treatment were carried out essentially as described above. Immune complexes were separated on SDS-PAGE, transferred to PVDF membrane, and analysed by autoradiography and immunoblotting.

Mass spectrometry. COS7 cells expressing myc-tagged wild-type human DISC1 were treated with or without okadaic acid (0.5 μM ; Calbiochem). Cells were lysed in RIPA buffer (150 mM NaCl, 50 mM Tris HCl, pH 7.5, 1% Nonidet P-40, 0.1% SDS, 0.5% sodium deoxycholate) containing protease inhibitor mixture (Roche Applied Sciences) and phosphatase inhibitor cocktail (Sigma). The solubilized proteins were immunoprecipitated by using an anti-myc antibody (Santa Cruz).

The precipitated proteins were separated by SDS-PAGE under non-reducing conditions and visualized by colloidal Coomassie staining. Gel-purified DISC1 protein was digested with trypsin and sequence analysis was performed by using microcapillary reverse-phase HPLC nano-electrospray tandem mass spectrometry on a Finnigan LTQ quadrupole ion trap mass spectrometer in the mass spectrometry facility at JHU. Identified sequences were confirmed by manually inspecting CID spectra. Protein identifications were considered significant if at least two individual peptide Mascot scores were above the Mascot calculated threshold.

In vitro kinase assays. GST-tagged human C-terminal (amino acids 598–854), human N-terminal (amino acids 1–348), mouse C-terminal (amino acids 594–853), and their site-directed mutants were generated in *Escherichia coli*, and purified. These recombinant proteins were incubated with purified recombinant PKA (New England Biolabs) or CDK5/p35 (Sigma) for 30 min at 30 °C. Reactions were supplemented with 10 $\mu\text{Ci ml}^{-1}$ of [γ -³²P]ATP (PerkinElmer Life Sciences) and 1 mM MgATP. The phosphorylation reactions were terminated with SDS sample buffer, and samples were analysed by SDS-PAGE followed by autoradiography and immunoblotting.

Fractionation by sucrose gradient. HEK293T cells 3 days after transfection with human DISC1 RNAi constructs together with A710- or E710-DISC1 expression construct were collected and fractionated by a discontinuous sucrose gradient, as described³⁸. In brief, cells were treated with 0.2 μM nocodazole for 1 h. Cells were washed in Tris-buffered saline (TBS), then trypsinized and homogenized in 0.1 \times TBS/8% sucrose. After centrifugation at 1,000g for 5 min, cells were resuspended in 0.1 \times TBS/8% sucrose followed by addition of lysis buffer (1 mM HEPES, pH 7.2, 0.5% Nonidet P-40, 0.5 mM MgCl₂, 0.1% β -mercaptoethanol, and protease inhibitor mixture, Roche Applied Sciences). The lysate was centrifuged at 2,500g for 10 min to remove swollen nuclei, chromatin aggregates, and unlysed cells. To the resulting supernatant fraction HEPES buffer and DNaseI were added to a final concentration of 10 μM and 1 $\mu\text{g ml}^{-1}$, respectively, and incubated on ice for 30 min. The mixture was loaded onto a discontinuous sucrose gradient consisting of 70, 50, and 40% solutions from the bottom, respectively, and centrifuged at 40,000g for 1 h. Fractions were collected from the bottom and stored at -80°C for further analysis.

In vitro binding assays. Maltose-binding protein fused (MBP-DISC1) recombinant proteins were generated as described³³. GST-GSK3 β and GST-BBS1 were

purchased from SignalChem and Abnova, respectively. Proteins and an antibody against GST were incubated in 150 mM NaCl, 50 mM Tris-HCl, pH 7.5, 1% Triton X-100, 0.1 mg ml⁻¹ BSA, and protease inhibitor mixture (Roche Applied Sciences) overnight at 4 °C. MBP-DISC1 bound to GST-GSK3 β or GST-BBS1 was precipitated with Sepharose beads. The protein precipitates were analysed with SDS-PAGE, followed by western blotting with an antibody against DISC1 (D27).

Statistical analyses. Optical density of immunoreactivity in western blotting was obtained using Image J software.

For determination of the statistical significance between two groups, either the Student's *t*-test (equal variances) or the modified Welch *t*-test (unequal variances) was used. Result of the F test was used to decide which test was appropriate. To compare three or more groups, one-way ANOVA followed by Bonferroni post hoc for multiple comparisons was used. Probability values (*P* values) < 0.05 were considered to be statistically significant (**P* < 0.05, ***P* < 0.01, ****P* < 0.001). Values depicted are means \pm s.e.m.

31. Duan, X. *et al.* Disrupted-In-Schizophrenia 1 regulates integration of newly generated neurons in the adult brain. *Cell* **130**, 1146–1158 (2007).
32. Niwa, M. *et al.* Knockdown of DISC1 by in utero gene transfer disturbs postnatal dopaminergic maturation in the frontal cortex and leads to adult behavioral deficits. *Neuron* **65**, 480–489 (2010).
33. Hayashi-Takagi, A. *et al.* Disrupted-in-Schizophrenia 1 (DISC1) regulates spines of the glutamate synapse via Rac1. *Nature Neurosci.* **13**, 327–332 (2010).
34. Wang, X., Qiu, R., Tsark, W. & Lu, Q. Rapid promoter analysis in developing mouse brain and genetic labeling of young neurons by doublecortin-DsRed-express. *J. Neurosci. Res.* **85**, 3567–3573 (2007).
35. Tan, P. L. *et al.* Loss of Bardet-Biedl syndrome proteins causes defects in peripheral sensory innervation and function. *Proc. Natl Acad. Sci. USA* **104**, 17524–17529 (2007).
36. Ishizuka, K. *et al.* Evidence that many of the DISC1 isoforms in C57BL/6J mice are also expressed in 129S6/SvEv mice. *Mol. Psychiatry* **12**, 897–899 (2007).
37. Regad, T., Bellodi, C., Nicotera, P. & Salomoni, P. The tumor suppressor Pml regulates cell fate in the developing neocortex. *Nature Neurosci.* **12**, 132–140 (2009).
38. Moudjou, M., Bordes, N., Paintrand, M. & Bornens, M. γ -Tubulin in mammalian cells: the centrosomal and the cytosolic forms. *J. Cell Sci.* **109**, 875–887 (1996).

Ancestral polyploidy in seed plants and angiosperms

Yuannian Jiao^{1,2}, Norman J. Wickett², Saravanaraj Ayyampalayam³, André S. Chanderbali⁴, Lena Landherr², Paula E. Ralph², Lynn P. Tomsho⁵, Yi Hu², Haiying Liang⁶, Pamela S. Soltis⁷, Douglas E. Soltis⁴, Sandra W. Clifton⁸, Scott E. Schlarbaum⁹, Stephan C. Schuster⁵, Hong Ma^{1,2,10,11}, Jim Leebens-Mack³ & Claude W. dePamphilis^{1,2}

Whole-genome duplication (WGD), or polyploidy, followed by gene loss and diploidization has long been recognized as an important evolutionary force in animals, fungi and other organisms^{1–3}, especially plants. The success of angiosperms has been attributed, in part, to innovations associated with gene or whole-genome duplications^{4–6}, but evidence for proposed ancient genome duplications pre-dating the divergence of monocots and eudicots remains equivocal in analyses of conserved gene order. Here we use comprehensive phylogenomic analyses of sequenced plant genomes and more than 12.6 million new expressed-sequence-tag sequences from phylogenetically pivotal lineages to elucidate two groups of ancient gene duplications—one in the common ancestor of extant seed plants and the other in the common ancestor of extant angiosperms. Gene duplication events were intensely concentrated around 319 and 192 million years ago, implicating two WGDs in ancestral lineages shortly before the diversification of extant seed plants and extant angiosperms, respectively. Significantly, these ancestral WGDs resulted in the diversification of regulatory genes important to seed and flower development, suggesting that they were involved in major innovations that ultimately contributed to the rise and eventual dominance of seed plants and angiosperms.

Angiosperms are by far the largest group of land plants, with more than 300,000 living species. Significantly, most flowering plant lineages reflect one or more rounds of ancient polyploidy. For example, extensive analyses of the complete genome sequence of *Arabidopsis thaliana* support two recent WGDs (named α and β) within the crucifer (Brassicaceae) lineage and one triplication event (γ) that is probably shared by all core eudicots^{7–13}. The *Populus trichocarpa* genome shows evidence of the core eudicot triplication as well as a more recent WGD¹⁴. Two polyploidy events in monocots (ρ and σ) have been inferred to have pre-dated the diversification of cereal grains and other grasses¹⁵ (Poaceae). Several studies have hinted that an ancient WGD event occurred even earlier in angiosperm evolution^{4,5,10,16}. However, the existence and timing of these ancient events, and their long-term impact, remain uncertain.

Here we use a rigorous phylogenomic approach (Supplementary Fig. 1; details in Supplementary Methods) to test the hypothesis that one or more ancient genome duplications occurred before the divergence of monocots and eudicots. By mapping the duplication events onto phylogenetic trees, we determine whether the paralogues were duplicated before or after a given speciation event^{8,17} (Fig. 1a). Although individual genes might be lost in some phylogenies, a broad picture can be drawn from simultaneous consideration of many or all gene families.

We used species with completely sequenced genomes (Supplementary Table 1; two monocots (*Oryza sativa* and *Sorghum bicolor*) and five eudicots (*A. thaliana*, *Carica papaya*, *P. trichocarpa*, *Cucumis sativus*

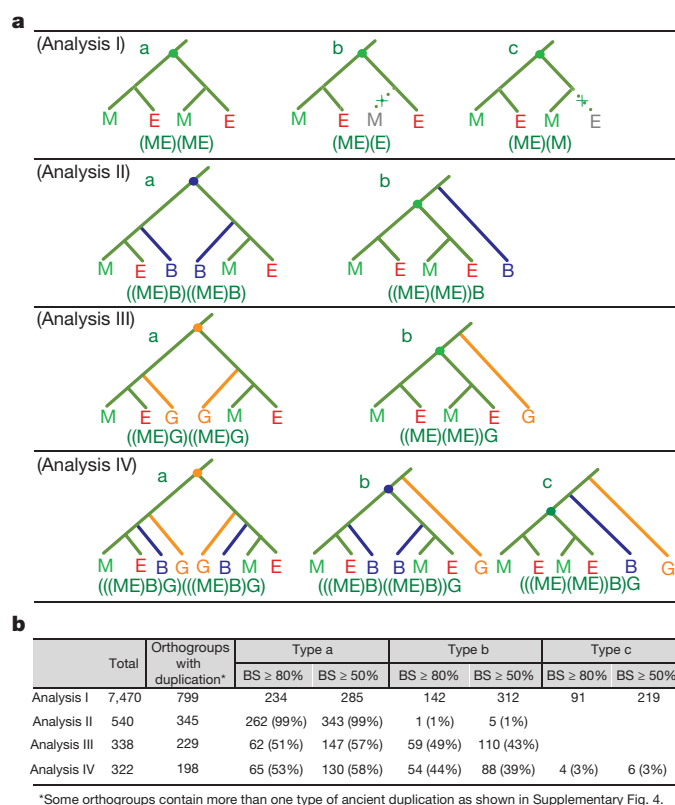


Figure 1 | Hypothetical tree topologies and summary of orthogroups consistent with ancient gene duplications before the split of monocots and eudicots. **a**, Analysis I: three examples of phylogenetic trees showing the patterns of retention or loss of paralogues: (a) both of the paralogues are retained in monocots and eudicots; (b) one of the paralogues was lost in monocots; (c) one of the paralogues was lost in eudicots. Analysis II: orthologues from basal angiosperms were added to core-orthogroups to refine the timing of ancient gene duplications in angiosperms: (a) gene duplication shared across all angiosperms; (b) gene duplication shared only by monocots and eudicots. Analysis III: orthologues from gymnosperms were added to core-orthogroups to place shared gene duplications before (a) and/or after (b) the split of extant gymnosperms and angiosperms. Analysis IV: three different topologies consistent with the timing of duplications shared by seed plants (a), angiosperms (b) and monocots + eudicots (c) when we expanded core-orthogroups with additional orthologues from both basal angiosperms and gymnosperms. M, monocots; E, eudicots; B, basal angiosperms; G, gymnosperms. Exemplar trees in analyses II, III and IV illustrate expected patterns with all branches retained. Observed topologies typically had partial gene losses similar to analyses Ib and Ic. **b**, Summary of orthogroups showing different types of duplications corresponding to proposed topologies inferred from orthogroup trees.

¹Intercollege Graduate Degree Program in Plant Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. ²Department of Biology, Institute of Molecular Evolutionary Genetics, and the Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. ³Department of Plant Biology, University of Georgia, Athens, Georgia 30602, USA. ⁴Department of Biology, University of Florida, Gainesville, Florida 32611, USA. ⁵Center for Comparative Genomics, Center for Infectious Disease Dynamics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. ⁶Department of Genetics and Biochemistry, Clemson University, Clemson, South Carolina 29634, USA. ⁷Florida Museum of Natural History, University of Florida, Gainesville, Florida 32611, USA. ⁸The Genome Center at Washington University, Saint Louis, Missouri 63108, USA. ⁹Department of Forestry, Wildlife & Fisheries, Institute of Agriculture, The University of Tennessee, Knoxville, Tennessee 37996, USA. ¹⁰State Key Laboratory of Genetic Engineering, School of Life Sciences, Institute of Plant Biology, Center for Evolutionary Biology, Fudan University, Shanghai 200433, China. ¹¹Institute of Biomedical Sciences, Fudan University, Shanghai 200433, China.

and *Vitis vinifera*) to construct gene families or subfamilies. One lycophyte (*Selaginella moellendorffii*) and one moss (*Physcomitrella patens*) were used as outgroups when dating gene duplications and potential WGDs that occurred before the monocot–eudicot divergence. In total, 77.03% of all protein-coding genes in the sequenced genomes were grouped in 31,433 multigene ‘core-orthogroups’. We define orthogroups as clusters of homologous genes that derive from a single gene in the common ancestor of the focal taxa, and refer to orthogroups for the nine sequenced genomes as core-orthogroups. Of these, 7,470 core-orthogroups contain at least one monocot, one eudicot, and one *Selaginella* and/or *Physcomitrella* sequence. These core-orthogroups were used in our investigation of duplication events predating the divergence of monocots and eudicots.

We queried maximum-likelihood trees (MLTs) for each core-orthogroup for topologies indicative of shared duplications (Fig. 1a, analysis I). We filtered our gene trees (Supplementary Methods), requiring that at least one of the seven core species retained both paralogues following the inferred gene duplication event in a common monocot–eudicot ancestor (see Supplementary Data 1 for a list of orthogroups). For example, the MLT for orthogroup 1711 (DEAD-box RNA helicase) contained duplicate genes in both monocots and eudicots whereas the MLTs for orthogroup 2312 (spermidine synthase) and orthogroup 396 (function unknown) showed that either one of the monocot or eudicot paralogues was lost after the divergence of monocots and eudicots (see exemplar trees in Supplementary Figs 2a, 3a and 4). On the basis of this conservative criterion, we identified a large number of core-orthogroups with shared duplication of monocots and eudicots (829 duplications in 799 core-orthogroups with bootstrap support (BS) greater than or equal to 50%; 474 duplications in 451 core-orthogroups with BS $\geq 80\%$; Supplementary Data 2). These duplications occurred before the γ triplication^{9,13} (which may be restricted to eudicots). As expected^{9,13}, many younger duplications within the sampled eudicot lineages were also observed on these trees (1,146 orthogroups surviving at least one eudicot-wide triplication (γ)), but for this study we focused on ancient duplications that occurred before the divergence of monocots and eudicots.

Additional homologues from basal angiosperms (*Aristolochia*, *Liriodendron*, *Nuphar* and *Amborella*; Supplementary Table 2) and gymnosperms (*Pinus*, *Picea*, *Zamia*, *Cryptomeria* and others; Supplementary Table 2) were added to 799 core-orthogroups to form expanded orthogroups¹⁸. These phylogenetically critical lineages increase gene sampling and provide better resolution of the timing of ancient duplications. By ‘basal angiosperms’ we mean the earliest-branching lineages of flowering plants that arose before the separation of monocots and eudicots. Before re-estimating gene trees for the expanded orthogroups, we added another quality control step to remove short or highly divergent unigenes (sequences produced from assembly of expressed-sequence-tag data sets; Supplementary Methods). After filtering, there remained 540 and 338 orthogroups with unigenes sampled from basal angiosperms and gymnosperms, respectively. Among these, 322 orthogroups contained unigenes from both basal angiosperms and gymnosperms (Fig. 1b).

For the 540 orthogroups with unigenes from basal angiosperms, the number of trees in which we identified an ancestral duplication before the origin of angiosperms¹⁹ (Fig. 1a, analysis IIa) greatly exceeded the number in which we identified a shared duplication after the origin of angiosperms (Fig. 1a, analysis IIb). Inference of a duplication pre-dating the diversification of basal angiosperms (ancestral angiosperm duplication) was supported by 262 (BS $\geq 80\%$) or 343 (BS $\geq 50\%$) orthogroups, whereas only one (BS $\geq 80\%$) or five (BS $\geq 50\%$) orthogroups supported inference of a gene duplication just after the origin of the angiosperm crown group (Fig. 1b, analysis II). We also found only five orthogroups with a surviving duplication shared with some, but not all, sampled basal angiosperms. Although basal angiosperms are a grade (and not a clade), we represent them with a single line in Fig. 1a because the duplication signal is inclusive of all basal angiosperms.

Additional analyses of 338 orthogroups populated with unigenes of gymnosperms identified 62 (BS $\geq 80\%$) or 147 (BS $\geq 50\%$) trees containing a seed-plant-wide gene duplication and 59 (BS $\geq 80\%$) or 110 (BS $\geq 50\%$) trees with a later duplication shared only by angiosperms (Fig. 1b, analysis III). In addition, analyses of the 322 orthogroups expanded with orthologues from both basal angiosperms and gymnosperms also detected similar signals of the two ancient shared duplications: 65 (BS $\geq 80\%$) or 130 (BS $\geq 50\%$) trees showing an ancestral seed plant duplication (see exemplar tree in Supplementary Fig. 2b), and 54 (BS $\geq 80\%$) or 88 (BS $\geq 50\%$) trees supporting an ancestral angiosperm duplication (Supplementary Fig. 3b and Fig. 1b, analysis IV).

In summary, our conservative filtering procedure identified 799 trees with topologies suitable for testing hypotheses concerning the presence of ancient duplications. These trees provided overwhelming support for the presence of two groups of duplications, one in the common ancestor of all angiosperms and the other in the common ancestor of all seed plants. Several mechanisms could explain the concerted patterns of gene duplication revealed in the gene trees, including WGD or multiple segmental or chromosomal duplications. The most parsimonious interpretation of the existing data is ancient WGD. We performed divergence time analyses to investigate this hypothesis further.

If the proposed WGDs were real, the estimated dates for gene duplication events in independent gene trees would be expected to be similar. Alternatively, if the duplications were unrelated (that is, a collection of independent events), a uniform distribution of duplication times within the intervals between the origins of gymnosperms and angiosperms would be expected for the ancestral angiosperm duplicates or on the branch leading to seed plants for the ancestral seed plant duplicates. We calibrated 799 core-orthogroups supporting (BS $\geq 50\%$) ancient duplications before the separation of monocots and eudicots from analysis I and estimated the divergence times of 860 nodes in 774 core-orthogroups using the program R8S (Supplementary Methods).

We then analysed the distribution of the inferred duplication times using a Bayesian method that assigned divergence time estimates to classes specified by a mixture model²⁰. The distribution of duplication times was bimodal, with peaks 192 ± 2 (95% confidence interval) and 319 ± 3 million years (Myr) ago. Dates were clustered in two relatively short time intervals, suggesting that these duplications were not uniformly distributed (Fig. 2a). Furthermore, we also analysed the 499

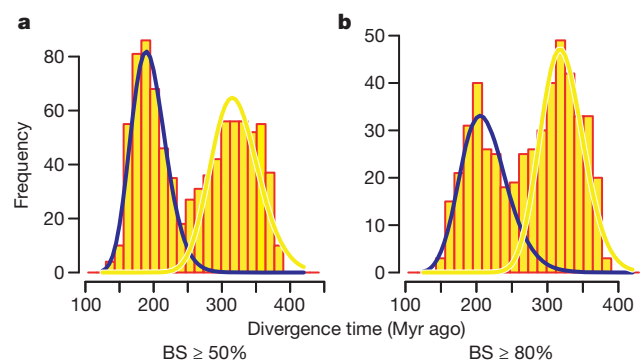


Figure 2 | Age distribution of ancient duplications shared by monocots and eudicots. **a**, The inferred divergence times for 866 ancestral duplication nodes in 779 core-orthogroups (BS $\geq 50\%$) were analysed by EMMIX to determine whether these duplications occurred randomly over time or within some small time frame. Each component is written as ‘colour/mean molecular timing/proportion’ where ‘colour’ is the component (curve) colour and ‘proportion’ is the percentage of duplication nodes assigned to the identified component. There are two statistically significant components: blue/192 (Myr ago)/0.48 and yellow/319/0.52. **b**, When we required the bootstrap support of the monocot + eudicot duplication to be greater than or equal to 80%, there were 504 nodes in 439 core-orthogroups for analysis of the inferred divergence times by EMMIX. Two statistically significant components were identified: blue/210/0.43 and yellow/321/0.57.

nodes with ancient duplications in 435 orthogroups with $BS \geq 80\%$ (Fig. 2b) and found a similar distribution pattern (two components: 210 ± 4 and 321 ± 4 Myr ago).

We then examined the age distribution of ancient duplications restricted only to orthogroups in analysis III that had been populated with nearly full-length gymnosperm unigenes. Among the 338 orthogroups with inferred absolute dates, there are 110 ($BS \geq 50\%$; 59 with $BS \geq 80\%$) that place a duplication on the angiosperm branch after divergence from gymnosperms. The distribution of duplication times inferred from these orthogroups showed one significant peak (234 ± 9 or 236 ± 9 Myr ago; Supplementary Fig. 5a, b). The most recent common ancestor of extant angiosperms existed has been dated to 130–190 Myr ago^{19,21}. Therefore, the identified duplication event occurred before the radiation of extant angiosperms, which agrees with the results from phylogenetic analysis (Fig. 1b, analysis II). An additional analysis was restricted to those 147 ($BS \geq 50\%$) or 62 ($BS \geq 80\%$) orthogroups (Fig. 1b, analysis IIIa) that contained a seed-plant-wide duplication based on phylogenetic analysis. The mixture model analysis identified only one significant component for the distribution of duplication times (349 ± 3 or 347 ± 4 Myr ago; Supplementary Fig. 5c, d), which was older than the ancestral node for extant seed plants²² (~310 Myr ago). Thus, both molecular dating and phylogenetic analyses support another ancient genome-wide duplication shared by all extant seed plants (Fig. 3). Distributions of synonymous site divergence for duplicated genes and synteny analyses also support this conclusion (Supplementary Discussion).

Gene duplication provides raw genetic material for the evolution of functional novelty. WGD in ancient seed plants would have generated duplicate copies of every gene, some of which could have had crucial roles in the origin of phenotypic novelty and, ultimately, in the origin and rapid diversification of the angiosperms. Although those genes retained as duplicates from the ancestral WGDs represent all functional categories, there is an overabundance of retained duplicate genes from several functional categories, including transferases and binding proteins, transcription factors and protein kinases (Supplementary Fig.

6 and Supplementary Data 3). These categories are significantly enriched for orthogroups surviving the monocot–eudicot duplication described in analysis I and for orthogroups surviving pre-angiosperm and/or pre-seed-plant duplications in analysis III. These results are consistent with patterns of gene retention following the more recent WGDs in the *Arabidopsis* lineage (ref. 23 and references therein), and WGD in vertebrates²⁴, supporting the interpretation that the concurrent duplications observed here are products of WGD. Taken together, these patterns suggest that the tendency for some types of gene duplicates to be retained following polyploidy has been a common feature of the post-WGD diploidization process throughout the evolutionary history of plants.

One subset of duplicated genes that could have contributed to ancient seed plant and angiosperm innovations includes those that have special roles in reproduction and flower development. In this study, we identified 35 orthogroups involved in flower developmental pathways with at least one ancient duplication event before the divergence of monocots and eudicots (Supplementary Table 3). For example, orthogroup 361 (containing *Arabidopsis* *PHYTOCHROME* genes), which includes regulators of flowering time²⁵ and seed germination²⁶, retained duplicate genes following two putative WGDs pre-dating the origin of angiosperms and seed plants, respectively, consistent with a published phylogeny for the *PHYTOCHROME* gene family²⁷. Other published gene family phylogenies also suggested common patterns of gene duplication, hinting at the genome-scale duplications seen here. For example, *TIR1/AFB* has experienced an ancient duplication before the diversification of extant angiosperms²⁸. Phylogenetic analyses of the *ZINC FINGER HOMEBOX (ZHD)* family²⁹, the *HD-ZIP III* gene family³⁰, and MADS-box genes (Supplementary Discussion) show duplication patterns consistent with WGDs pre-dating the origin of angiosperms and seed plants. Hence, these previous studies of individual genes or gene families bolster our conclusions based on a genome-wide survey of thousands of genes, and identify some of the many genes derived from these duplications that could potentially have had important roles in seed plant and angiosperm evolution.

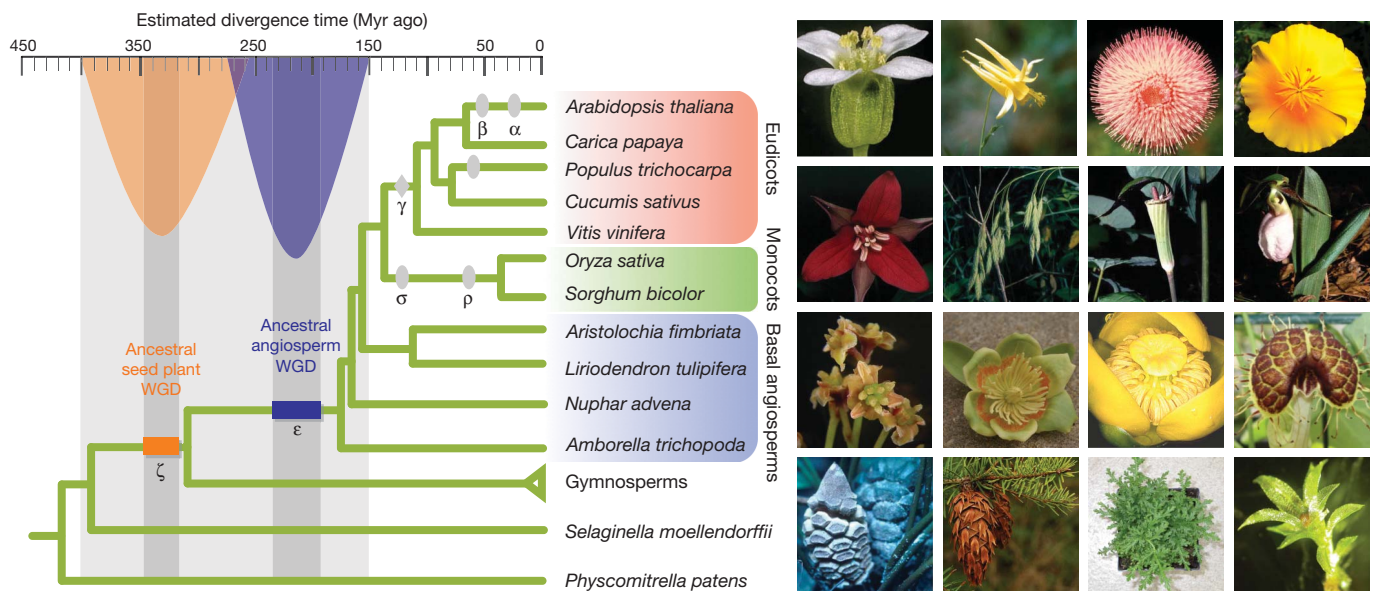


Figure 3 | Ancestral polyploidy events in seed plants and angiosperms. Two ancestral duplications identified by integration of phylogenomic evidence and molecular time clock for land plant evolution. Ovals indicate the generally accepted genome duplications identified in sequenced genomes (see text). The diamond refers to the triplication event probably shared by all core eudicots. Horizontal bars denote confidence regions for ancestral seed plant WGD and ancestral angiosperm WGD, and are drawn to reflect upper and lower bounds of mean estimates from Fig. 2 (more orthogroups) and Supplementary Fig. 5 (more taxa). The photographs provide examples of the reproductive diversity of

eudicots (top row, left to right: *Arabidopsis thaliana*, *Aquilegia chrysantha*, *Cirsium pumilum*, *Eschscholzia californica*), monocots (second row, left to right: *Trillium erectum*, *Bromus kalmii*, *Arisaema triphyllum*, *Cypripedium acaule*), basal angiosperms (third row, left to right: *Amborella trichopoda*, *Liriodendron tulipifera*, *Nuphar advena*, *Aristolochia fimbriata*), gymnosperms (fourth row, first and second from left: *Zamia vazequezii*, *Pseudotsuga menziesii*) and the outgroups *Selaginella moellendorffii* (vegetative; fourth row, third from left) and *Physcomitrella patens* (fourth row, right). See Supplementary Table 4 for photo credits.

METHODS SUMMARY

Phylogenetic analysis. We used the OrthoMCL method to construct a set of core-orthogroups. All orthogroup amino-acid alignments were generated with MUSCLE and then trimmed by removing poorly aligned regions using TRIMAL 1.2. Additional sorted unigene sequences for the core-orthogroups (retrieved with HaMSTR) were aligned at the amino-acid level into the existing nine species' full alignments (before trimming) using CLUSTALX 1.8. After trimming, each unigene sequence was checked and removed from the alignment if the sequence contained less than 70% alignment length. Corresponding DNA sequences were then forced onto the amino-acid alignment using custom Perl scripts and used for subsequent phylogenetic analysis. Maximum-likelihood analyses were conducted using RAXML, version 7.2.1, searching for the best MLT with the GTRGAMMA model, which represents an acceptable trade-off between speed and accuracy (RAXML 7.0.4 manual).

Molecular dating analyses and 95% confidence intervals. The divergence time of the two paralogous clades derived from each duplication was estimated from the best maximum-likelihood topologies under the assumption of a relaxed molecular clock by applying a semi-parametric penalized likelihood approach using a truncated Newton optimization algorithm as implemented in the program R8S. The smoothing parameter was determined by cross-validation. Dating constraints are described in Methods. The EMMIX software package was used to fit a mixture model of multivariate normal or *t*-distributed components to a given data set. For each significant component identified by EMMIX, the 95% confidence interval of the mean date estimate was then calculated.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 August 2010; accepted 10 February 2011.

Published online 10 April 2011.

- Ohno, S. *Evolution by Gene Duplication* (Springer, 1970).
- Lynch, M. *The Origins of Genome Architecture* (Sinauer, 2007).
- Edger, P. P. & Pires, J. C. Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res.* **17**, 699–717 (2009).
- De Bodt, S., Maere, S. & Van de Peer, Y. Genome duplication and the origin of angiosperms. *Trends Ecol. Evol.* **20**, 591–597 (2005).
- Soltis, D. E., Bell, C. D., Kim, S. & Soltis, P. S. Origin and early evolution of angiosperms. *Ann. NY Acad. Sci.* **1133**, 3–25 (2008).
- Fawcett, J. A., Maere, S. & Van de Peer, Y. Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proc. Natl Acad. Sci. USA* **106**, 5737–5742 (2009).
- Lyons, E. *et al.* Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: CoGe with rodents. *Plant Physiol.* **148**, 1772–1781 (2008).
- Bowers, J. E., Chapman, B. A., Rong, J. & Paterson, A. H. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**, 433–438 (2003).
- Jaillon, O. *et al.* The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463–467 (2007).
- Vision, T. J., Brown, D. G. & Tanksley, S. D. The origins of genomic duplications in *Arabidopsis*. *Science* **290**, 2114–2117 (2000).
- Barker, M. S., Vogel, H. & Schranz, M. E. Paleopolyploidy in the Brassicales: analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. *Genome Biol. Evol.* **1**, 391–399 (2009).
- Tang, H. *et al.* Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008).
- Tang, H. *et al.* Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* **18**, 1944–1954 (2008).
- Tuskan, G. A. *et al.* The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596–1604 (2006).
- Tang, H., Bowers, J. E., Wang, X. & Paterson, A. H. Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proc. Natl Acad. Sci. USA* **107**, 472–477 (2010).
- Cui, L. *et al.* Widespread genome duplications throughout the history of flowering plants. *Genome Res.* **16**, 738–749 (2006).
- Blomme, T. *et al.* The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol.* **7**, R43 (2006).
- Ebersberger, I., Strauss, S. & von Haeseler, A. HaMSTR: profile hidden Markov model based search for orthologs in ESTs. *BMC Evol. Biol.* **9**, 157 (2009).
- Moore, M. J., Bell, C. D., Soltis, P. S. & Soltis, D. E. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc. Natl Acad. Sci. USA* **104**, 19363–19368 (2007).
- McLachlan, G., Peel, D., Basford, K. E. & Adams, P. The EMMIX algorithm for the fitting of normal and *t*-components. *J. Stat. Softw.* **4**, i02 (1999).
- Bell, C. D., Soltis, D. E. & Soltis, P. S. The age of the angiosperms: a molecular timescale without a clock. *Evolution* **59**, 1245–1258 (2005).
- Schneider, H. *et al.* Ferns diversified in the shadow of angiosperms. *Nature* **428**, 553–557 (2004).
- Freeling, M. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* **60**, 433–453 (2009).
- Kassahn, K. S., Dang, V. T., Wilkins, S. J., Perkins, A. C. & Ragan, M. A. Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. *Genome Res.* **19**, 1404–1418 (2009).
- Devlin, P. F., Patel, S. R. & Whitelam, G. C. Phytochrome E influences internode elongation and flowering time in *Arabidopsis*. *Plant Cell* **10**, 1479–1487 (1998).
- Dechaine, J. M., Gardner, G. & Weinig, C. Phytochromes differentially regulate seed germination responses to light quality and temperature cues during seed maturation. *Plant Cell Environ.* **32**, 1297–1309 (2009).
- Mathews, S., Burleigh, J. G. & Donoghue, M. J. Adaptive evolution in the photosensory domain of phytochrome A in early angiosperms. *Mol. Biol. Evol.* **20**, 1087–1097 (2003).
- Parry, G. *et al.* Complex regulation of the *TIR1/AFB* family of auxin receptors. *Proc. Natl Acad. Sci. USA* **106**, 22540–22545 (2009).
- Hu, W., dePamphilis, C. W. & Ma, H. Phylogenetic analysis of the plant-specific zinc finger-homeobox and mini zinc finger gene families. *J. Integr. Plant Biol.* **50**, 1031–1045 (2008).
- Prigge, M. J. & Clark, S. E. Evolution of the class III HD-Zip gene family in land plants. *Evol. Dev.* **8**, 350–361 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported primarily by NSF Plant Genome Research Program (DEB 0638595, The Ancestral Angiosperm Genome Project) and in part by the Department of Biology and by the Huck Institutes of Life Sciences of the Pennsylvania State University. H.M. was also supported by funds from Fudan University. We thank J. Carlson, M. Frohlich, S. DiLoretto, L. Warg, S. Crutchfield, C. Johnson, N. Naznin, X. Zhou, J. Duarte, B. J. Bliss, J. Der and E. Wafala for help and discussion, D. Stevenson and C. Schultz for *Zamia* samples, J. McNeal, S. Kim and M. Axtell for photographs, and all the members of The Genome Center at Washington University production team, especially L. Fulton, K. Delehaunty and C. Fronick.

Author Contributions Y.J. and C.W.d. designed the study and Y.J. performed the principal data analyses. A.S.C., L.L., P.E.R., Y.H., S.E.S. and H.L. prepared tissues, RNAs, and/or libraries. S.W.C., L.P.T. and S.C.S. generated sequence data. S.A. and J.L.-M. performed the Ancestral Angiosperm Genome Project transcriptome assemblies and MAGIC database construction. Y.J. and C.W.d. drafted the manuscript, and N.J.W., A.S.C., L.L., P.E.R., P.S.S., D.E.S., H.M. and J.L.-M. contributed to the planning and discussion of the research and the editing of the manuscript. All authors contributed to and approved the final manuscript.

Author Information Alignments and phylogenetic trees have been deposited in Dryad with package identifier doi:10.5061/dryad.8546. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.W.d. (cw3d@psu.edu).

METHODS

Detection of ancient WGD events. Several methodologies have been proposed and widely used to detect the signature of genome duplication. Identification of large syntenic blocks of genes within genomes provides strong evidence to support genome duplication^{7,8}. The timing of WGDs is inferred through cross-species genome comparisons, but extensive genome rearrangements and gene loss reduce the size of syntenic blocks over time and obscure identification of ancient pre- γ WGD^{31,32}. Another approach is to estimate the age distribution of paralogous gene pairs, where synonymous site divergence (K_s) or non-synonymous site divergence (K_a) is used as a proxy for the age of the duplication event^{4,10,16,33}. However, this method may be confounded by excessive gene loss, concentration of duplicate pair estimates on more recent nodes, saturation of K_s between older paralogue pairs, and molecular rate heterogeneity among lineages, gene families or even genes. For example, the β and γ WGDs inferred in analyses of syntenic blocks were not evident in a K_s plot for *Arabidopsis* paralogue pairs^{13,33,34}. Therefore, both methods present challenges to inferring ancient genome duplications that may have occurred close to or well before the origin of angiosperms. For this reason, we used phylogenomic analyses to identify ancient gene duplications that occurred before monocots and dicots, and evaluated their phylogenetic timing and estimated age to identify whether there were temporal concentrations of gene duplications (Supplementary Fig. 1).

Phylogenetic analysis. The OrthoMCL method³⁵ was used to construct a set of core-orthogroups based on protein similarity graphs. This approach has been shown to yield fewer false positives than other methods³⁶, which is critical for this study. If genes from outside the core-orthogroup in question (false positives) are included in the analysis, the core-orthogroup could be incorrectly scored as retaining ancient duplicates. All orthogroup amino-acid alignments were generated with MUSCLE using default parameters³⁷. The multiple sequence alignments were trimmed by removing poorly aligned regions using TRIMAL 1.2 with the option 'automated1'³⁸. Additional sorted unigene sequences for the core-orthogroups (retrieved with HamStr) were aligned at the amino-acid level into the existing 11 species' full alignments (before trimming) using CLUSTALX 1.8³⁹. After trimming, each unigene sequence was checked and removed from the alignment if the sequence contained less than 70% alignment coverage. Corresponding DNA sequences were then forced onto the amino-acid alignment using custom Perl scripts and used for subsequent phylogenetic analysis. Maximum-likelihood analyses were conducted using RAXML, version 7.2.1^{40,41}, invoking a rapid bootstrap (100 replicates) analysis and search for the best-scoring MLT with the general time-reversible model of DNA sequence evolution with gamma-distributed rate heterogeneity (the GTRGAMMA model, which represents an acceptable trade-off between speed and accuracy; RAXML 7.0.4 manual) in a single program run. Alignments and phylogenetic trees are deposited at <http://dx.doi.org/10.5061/dryad.8546>, and Perl scripts are available on request from C.W.d.

Scoring gene duplications. By carefully interpreting all of the trees, duplication events were identified in rooted trees using *Physcomitrella* genes (or *Selaginella* if there were no *Physcomitrella* genes in the orthogroup) as outgroup sequences. Three relevant bootstrap values were taken into account when evaluating support for a particular duplication. For example, given a topology of (((M1E1)bootstrap1,(M2E2)bootstrap2)bootstrap3), bootstrap1 and bootstrap2 are the bootstrap values supporting the M1E1 clade and the M2E2 clade, respectively, and bootstrap3 is the bootstrap value supporting the large clade including M1E1 and M2E2. A monocot-eudicot duplication supported by 50% (or 80%) means that bootstrap3 and at least one of the bootstrap1 and bootstrap2 values are greater than or equal to 50% (or 80%). When basal angiosperm and/or gymnosperm genes were added, bootstrap1 and bootstrap2 were evaluated for nodes subtending ME + B (Fig. 1a), whereas bootstrap3 was evaluated for the node subtending the large clade including the angiosperm-wide or seed-plant-wide duplications.

Gene tree estimation may be susceptible to long-branch attraction, particularly with sparse taxon sampling (that is, sparse gene sampling in the gene tree context) or when there is mis-specification of the model of molecular evolution used for phylogenetic reconstruction^{42,43}, leading to erroneous conclusions of topology. For example, an orthogroup with the phylogenetic pattern ((*Oryza*, *Populus*)(*Arabidopsis*)) is consistent with a gene duplication shared by monocots and eudicots, with subsequent paralogous losses in both monocot and eudicot lineages (Fig. 1a, analysis 1b). Alternatively, it is possible that the *Arabidopsis* gene was especially divergent and therefore was placed as sister to the *Oryza*-*Populus* pair owing to long-branch attraction. Distinguishing between these alternative explanations can be facilitated by increased gene sampling to split long branches⁴³. Moreover, inference of gene duplication may be ambiguous if all taxa are represented by a single gene in a given gene tree (as in the example above). With these considerations in mind, we filtered our gene trees, requiring that at least one of the seven core species has retained both paralogues following the inferred gene duplication event in a common monocot-eudicot ancestor. Therefore, an example of the smallest possible gene tree with a monocot-eudicot duplication would be (((*Oryza*, *Vitis*)(*Vitis*))*Selaginella*). On the basis of these criteria, we scored each orthogroup with

or without ancient duplications, and counted the total number of orthogroups supporting each hypothesis illustrated in Fig. 1a. Supplementary Data 2 details the number of duplication of each type scored for every orthogroup.

Finite mixture models of genome duplications. To explore the timing of genome duplication events, the inferred distribution of divergence times was fitted to a mixture model comprising several component distributions in various proportions. The EMMIX software²⁰ can be used to fit a mixture model of multivariate normal or t -distributed components to a given data set (<http://www.maths.uq.edu.au/~gjm/emmix/emmix.html>). The mixed populations were modelled with one to four components. The EM algorithm was repeated 100 times with random starting values, as well as ten times with k -mean starting values. The best mixture model was identified using the Bayesian information criterion.

Molecular dating analyses and 95% confidence intervals. The best maximum-likelihood topology for the core-orthogroups or orthogroups was used for divergence time analyses. The divergence time of the two paralogous clades was estimated under the assumption of a relaxed molecular clock by applying a semi-parametric penalized likelihood approach using a truncated Newton optimization algorithm as implemented in the program R8S⁴⁴. The smoothing parameter was determined by cross-validation. We used the following dates in our estimation procedure: minimum age of 400 Myr and maximum age of 450 Myr for the divergence of *P. patens*⁴⁵, a fixed constraint age of 400 Myr for the divergence of *S. moellendorffii*⁴⁶, a minimum age of 309 Myr for crown-group seed plants⁴⁷ (this constraint was used only in analyses reported in Supplementary Fig. 5), a minimum age of 125 Myr for the divergence of monocots and eudicots⁴⁸, and a maximum age of 125 Myr for the origin of rosids⁴⁸. We required that trees pass both the cross-validation procedure and provide estimates of the age of the duplication node. The inferred divergence times were then analysed by EMMIX. For each significant component identified by EMMIX, the 95% confidence interval of the mean was then calculated.

Calculation of K_s . Paralogous pairs of sequences were identified from best reciprocal matches in all-by-all BLASTN searches. Only protein sequences more than 200 base pairs in length were used for K_s calculations. Translated sequences of unigenes generated by ESTSCAN were aligned using MUSCLE 3.6³⁷. Nucleotide sequences were then forced to fit the amino-acid alignments using PAL2NAL⁴⁹. The K_s (also known as D_s) values were calculated using a simplified version of the Goldman-Yang maximum-likelihood method⁵⁰ implemented in the 'codeml' package of PAML⁵¹. The K_s frequency in each interval size of 0.05 within the range [0, 3.0] was plotted.

Gene ontology enrichment for orthogroups with ancient duplication. Gene ontology (GO) annotations of orthogroups with early ancient duplications were compared with orthogroups that did not have such duplications, to test for enrichment of GO terms⁵². *Arabidopsis* GO slim terms were downloaded and assigned to orthogroups directly if the orthogroup included *Arabidopsis* genes. Otherwise, we searched representative InterPro domains using INTERPROSCAN⁵³. Then GO annotations were assigned to the orthogroups using InterPro2GO mapping. Subsequently, all GO annotations were mapped to GO slim categories using the 'map2slim' script. Finally, we evaluated statistical differences in enrichment of GO slim terms using agriGO by Fisher's exact test and the Yekutieli (false-discovery rate under dependency) multi-test adjustment method⁵⁴.

- Buggs, R. J. *et al.* Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids. *Heredity* **103**, 73–81 (2009).
- Vandepoele, K., Simillion, C. & Van de Peer, Y. Detecting the undetectable: uncovering duplicated segments in *Arabidopsis* by comparison with rice. *Trends Genet.* **18**, 606–608 (2002).
- Blanc, G. & Wolfe, K. H. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* **16**, 1667–1678 (2004).
- Van de Peer, Y., Fawcett, J. A., Proost, S., Sterck, L. & Vandepoele, K. The flowering world: a tale of duplications. *Trends Plant Sci.* **14**, 680–688 (2009).
- Li, L., Stoeckert, C. J. Jr & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
- Proost, S. *et al.* PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *Plant Cell* **21**, 3718–3731 (2009).
- Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
- Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
- Thompson, J. D., Gibson, T. J. & Higgins, D. G. Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinf.* 2.3.1–2.3.22 (2002).
- Stamatakis, A., Ludwig, T. & Meier, H. RAXML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21**, 456–463 (2005).
- Stamatakis, A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
- Felsenstein, J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401–410 (1978).

43. Hendy, M. D. & Penny, D. A framework for the quantitative study of evolutionary trees. *Syst. Zool.* **38**, 297–309 (1989).
44. Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003).
45. Rensing, S. A. *et al.* The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* **319**, 64–69 (2008).
46. Kenrick, P. & Crane, P. R. The origin and early evolution of plants on land. *Nature* **389**, 33–39 (1997).
47. Miller, C. N. J. Implications of fossil conifers for the phylogenetic relationships of living families. *Bot. Rev.* **65**, 239–277 (1999).
48. Doyle, J. A. & Hotton, C. L. in *Pollen and Spores: Patterns of Diversification* (eds Blackmore, S. & Barnes, S. H.) 169–195 (Clarendon, 1991).
49. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612 (2006).
50. Goldman, N. & Yang, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**, 725–736 (1994).
51. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
52. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature Genet.* **25**, 25–29 (2000).
53. Zdobnov, E. M. & Apweiler, R. InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848 (2001).
54. Du, Z., Zhou, X., Ling, Y., Zhang, Z. & Su, Z. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* **38**, W64–W70 (2010).

Suppression of lung adenocarcinoma progression by Nkx2-1

Monte M. Winslow^{1,2}, Talya L. Dayton¹, Roel G. W. Verhaak^{3,4}, Caroline Kim-Kiselak^{1,2}, Eric L. Snyder¹, David M. Feldser¹, Diana D. Hubbard^{3,4}, Michel J. DuPage¹, Charles A. Whittaker¹, Sebastian Hoersch¹, Stephanie Yoon¹, Denise Crowley¹, Roderick T. Bronson⁵, Derek Y. Chiang^{3,4,6}, Matthew Meyerson^{3,4} & Tyler Jacks^{1,2,7,8}

Despite the high prevalence and poor outcome of patients with metastatic lung cancer the mechanisms of tumour progression and metastasis remain largely uncharacterized. Here we modelled human lung adenocarcinoma, which frequently harbours activating point mutations in *KRAS*¹ and inactivation of the *p53* pathway², using conditional alleles in mice^{3–5}. Lentiviral-mediated somatic activation of oncogenic *Kras* and deletion of *p53* in the lung epithelial cells of *Kras*^{LSL-G12D/+}; *p53*^{flx/flx} mice initiates lung adenocarcinoma development⁴. Although tumours are initiated synchronously by defined genetic alterations, only a subset becomes malignant, indicating that disease progression requires additional alterations. Identification of the lentiviral integration sites allowed us to distinguish metastatic from non-metastatic tumours and determine the gene expression alterations that distinguish these tumour types. Cross-species analysis identified the NK2-related homeobox transcription factor Nkx2-1 (also called Ttf-1 or Titf1) as a candidate suppressor of malignant progression. In this mouse model, Nkx2-1 negativity is pathognomonic of high-grade poorly differentiated tumours. Gain- and loss-of-function experiments in cells derived from metastatic and non-metastatic tumours demonstrated that Nkx2-1 controls tumour differentiation and limits metastatic potential *in vivo*. Interrogation of Nkx2-1-regulated genes, analysis of tumours at defined developmental stages, and functional complementation experiments indicate that Nkx2-1 constrains tumours in part by repressing the embryonically restricted chromatin regulator *Hmga2*. Whereas focal amplification of *NKX2-1* in a fraction of human lung adenocarcinomas has focused attention on its oncogenic function^{6–9}, our data specifically link Nkx2-1 downregulation to loss of differentiation, enhanced tumour seeding ability and increased metastatic proclivity. Thus, the oncogenic and suppressive functions of Nkx2-1 in the same tumour type substantiate its role as a dual function lineage factor.

We induced lung tumours in mice harbouring a *loxP*-Stop-*loxP* *Kras*^{G12D} knockin allele and both alleles of *p53* flanked by *loxP* sites (*Kras*^{LSL-G12D/+}; *p53*^{flx/flx} mice) through intratracheal administration of lentiviral vectors that express Cre-recombinase (Lenti-Cre)¹⁰. A lentiviral dose was used such that each mouse developed between 5 and 20 lung tumours, lived 8–14 months after tumour initiation and developed macroscopic metastases to the draining lymph nodes, pleura, kidneys, heart, adrenal glands and liver (Supplementary Fig. 1). Because lentiviruses integrate stably into the genome, the integration site was a unique molecular identifier that unambiguously linked primary tumours to their related metastases (Fig. 1a). We used linker-mediated polymerase chain reaction (LM-PCR) to determine the genomic sequence directly 3' of the integrated lentiviral genome followed by a specific PCR for the lentiviral integration site (Fig. 1b). To have samples of sufficient quantity and purity for our analyses, we derived cell lines

from primary tumours and metastases. Cell lines were pure tumour cells as determined by recombination of the *p53*^{flx} alleles (data not shown). The clonal relationship of these cell lines was established using LM-PCR or Southern blot analysis for the lentiviral genome (Fig. 1c and data not shown). We termed cell lines derived from verified metastatic primary lung tumours as T_{Met}.

Gene expression profiling was performed on cell lines from 23 lung tumours and metastases (nine metastases, seven T_{Met} primary tumours and seven potentially non-metastatic primary tumours). Using

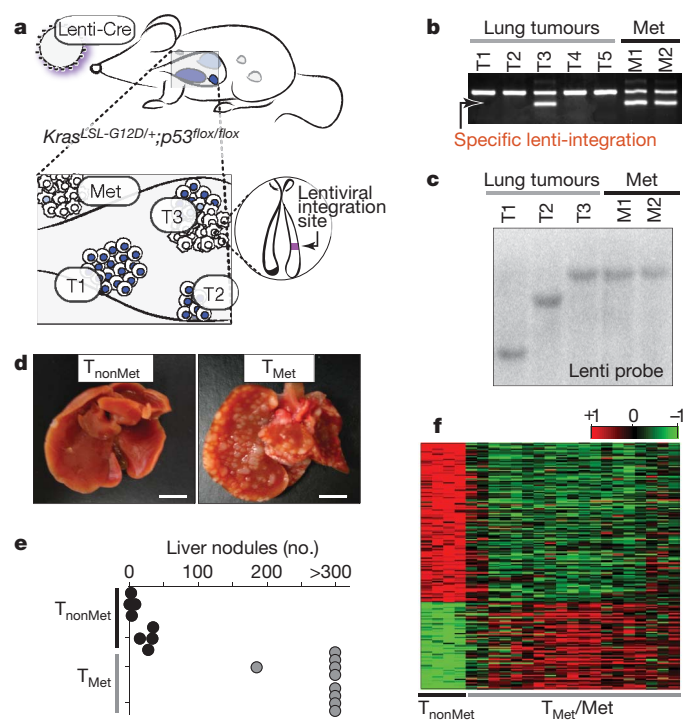


Figure 1 | A lentiviral vector-induced mouse model of lung adenocarcinoma identifies gene expression alterations during tumour progression. **a**, Infection of *Kras*^{LSL-G12D/+}; *p53*^{flx/flx} mice with Cre-expressing lentiviral vectors initiates lung adenocarcinoma. **b**, Linker-mediated PCR cloning of the lentiviral integration site in metastases (Met) allows specific PCR amplification of that lentiviral integration (lower band) to identify which primary tumour gave rise to the metastases. The top band is a control product. **c**, Southern blot on cell lines for the integrated lentiviral genome. **d**, Representative images of livers after intrasplenic transplantation of T_{nonMet} or T_{Met} cells. Scale bar = 0.5 cm. **e**, Quantification of liver nodules after intrasplenic injection of two T_{nonMet} and T_{Met} cell lines. **f**, Gene expression alterations (log₂) between T_{nonMet} and T_{Met}/Met samples.

¹David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Ludwig Center for Molecular Oncology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Dana-Farber Cancer Institute, Harvard University, Cambridge, Massachusetts 02115, USA. ⁴Broad Institute, Cambridge, Massachusetts 02142, USA. ⁵Department of Biomedical Sciences, Tufts University Veterinary School, North Grafton, Massachusetts 01536, USA. ⁶Department of Genetics, University of North Carolina, North Carolina 27599, USA. ⁷Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁸Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

unsupervised consensus clustering¹¹, we identified four cell lines from likely non-metastatic tumour samples that had highly concordant gene expression and were separate from all T_{Met} and metastasis (Met) samples (Supplementary Fig. 2). Therefore, we surmised that these could represent non-metastatic primary tumours and classified them as T_{nonMet} . These T_{nonMet} cell lines consistently formed fewer tumour nodules in the liver after intrasplenic injection despite equivalent proliferation rates in cell culture (Fig. 1d, e and Supplementary Fig. 2).

Significant gene expression alterations distinguished T_{nonMet} from T_{Met} and Met-derived cell lines (Fig. 1f and Supplementary Table 1), many of which were validated by quantitative reverse transcription PCR (qRT-PCR), flow cytometry and western blotting (data not shown). A gene expression signature generated by comparing T_{nonMet} to T_{Met} /Met samples predicted patient outcome in human lung adenocarcinoma gene expression data sets^{12,13}, indicating the possibility of evolutionarily conserved molecular mechanisms of tumour progression (Supplementary Fig. 2). Thus, we integrated mouse and human data by comparing the differences in expression between T_{nonMet} and T_{Met} /Met samples with the association of human gene expression and patient survival (Fig. 2a). Two genes were particularly notable from this analysis: the NK-related homeobox transcription factor *Nkx2-1* and the *Nkx2-1* target gene surfactant protein B (*Sftpb*; Fig. 2a). *Nkx2-1* regulates lung development and is expressed in type II pneumocytes and bronchiolar cells in the adult^{14–16}. *Nkx2-1* expression was >10-fold higher in T_{nonMet} samples, and higher *NKX2-1* expression in human tumours correlated with longer survival. Of note, *NKX2-1* is focally amplified in ~10% of human lung adenocarcinoma, with functional data supporting oncogenic activity in this setting^{6–9,17}. Conversely, most immunohistochemical analyses of *NKX2-1* in this disease suggest an association between *NKX2-1*-negative tumours and poor patient outcome^{17,18}. Thus, we focused on validating and characterizing the function of this transcription factor in suppressing tumour progression and metastasis.

We confirmed reduced *Nkx2-1* messenger RNA and protein in T_{Met} and Met cell lines without evidence of focal genomic loss of this region (Fig. 2b, Supplementary Fig. 4 and data not shown). *Nkx2-1* was

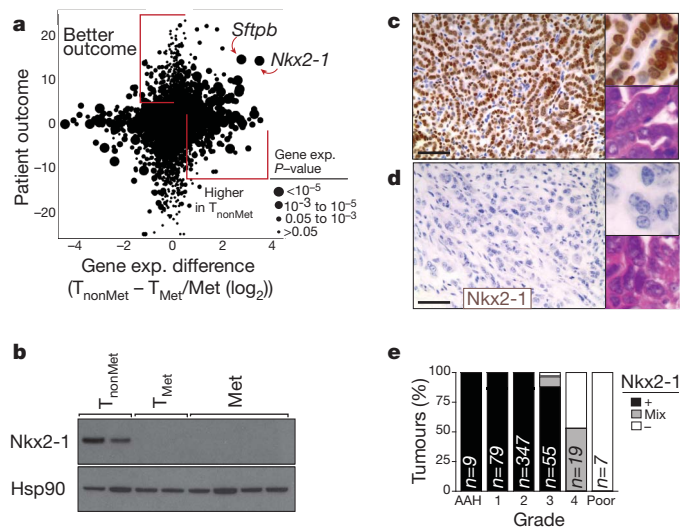


Figure 2 | Reduced *Nkx2-1* expression in advanced lung adenocarcinoma correlates with a less differentiated state. **a**, Cross-species analysis of human lung adenocarcinoma¹² patient outcome (likelihood ratio with the sign from correlation value) versus differential gene expression in murine T_{nonMet} cells. **b**, *Nkx2-1* protein is absent from T_{Met} and Met-derived cell lines. **c, d**, *Nkx2-1* expression is high in well differentiated adenomas and early murine adenocarcinoma (top) but is downregulated in moderately to poorly differentiated advanced carcinomas (bottom). Scale bar = 50 μm. Upper inset: *Nkx2-1* staining. Lower inset: haematoxylin and eosin staining. **e**, Quantification of *Nkx2-1* expression in murine lung tumours relative to tumour grade from most differentiated (atypical adenomatous hyperplasia (AAH)) to least differentiated (Poor)).

consistently downregulated in high-grade poorly differentiated tumours from our mouse model (Fig. 2c–e and Supplementary Fig. 3) as well as in advanced *Kras*^{G12D}-driven lung adenocarcinomas with *p53*^{R270H} or *p53*^{R172H} point mutations^{4,19}. Using our LM-PCR assay, we identified three primary lung tumours as metastatic on the basis of the presence of metastases with the same lentiviral integration site (Fig. 1b and data not shown). These primary tumours each contained poorly differentiated areas that were *Nkx2-1* negative (*Nkx2-1*[−]) (Supplementary Fig. 6). Interestingly, *Nkx2-1* expression was low/absent in almost all lymph node and distant macrometastases, although some micrometastases were *Nkx2-1*⁺ or *Nkx2-1*^{mixed} (Supplementary Fig. 3). Whether certain micrometastases were seeded by *Nkx2-1*⁺ cells or reverted to an *Nkx2-1*⁺ phenotype due to cues from their new environment is unknown.

In human lung adenocarcinoma^{12,13} the expression of *NKX2-1* correlated with a mouse T_{nonMet} gene expression signature (Supplementary Fig. 3). Additionally, the T_{nonMet} signature anti-correlated with an embryonic stem-cell signature, supporting the notion that T_{Met} /Met cells have transitioned to a less differentiated and more stem-like state²⁰ (Supplementary Fig. 3).

The correlative mouse and human data were consistent with *Nkx2-1* being either a marker or a functional regulator of tumour progression. *Nkx2-1* expression in a T_{Met} cell line (T_{Met} -*Nkx2-1* cells) greatly suppressed tumour formation after intravenous transplantation (Fig. 3a, b and Supplementary Fig. 5). Moreover, of the tumours that formed after injection of T_{Met} -*Nkx2-1* cells, many were either *Nkx2-1*[−] or *Nkx2-1*^{mixed} (Fig. 3c). In general, tumours that continued to express *Nkx2-1* were well differentiated, whereas *Nkx2-1*[−] tumours often displayed solid architecture or areas of poorly differentiated cells (Fig. 3d and Supplementary Fig. 5). Intrasplenic transplantation unveiled a similar diminution of tumour formation by T_{Met} -*Nkx2-1* cells (Supplementary Fig. 5). *Nkx2-1* expression did not alter proliferation or cell death in cell culture, or affect established tumour proliferation *in vivo* (Supplementary Fig. 5 and data not shown), but markedly reduced the ability of these cells to grow in anchorage-independent conditions and initiate tumours after subcutaneous transplantation (Fig. 3e and Supplementary Fig. 5).

To elucidate further the function of *Nkx2-1*, we knocked down *Nkx2-1* in T_{nonMet} cell lines using short hairpin RNA (shRNA; Supplementary Fig. 7). *Nkx2-1* knockdown allowed the formation of more liver nodules after intrasplenic injection and more lung nodules after intravenous transplantation (Fig. 3f). *Nkx2-1* knockdown did not alter proliferation or cell death in cell culture (Supplementary Fig. 7) but enhanced the cells' ability to form colonies under anchorage-independent conditions and tumours after subcutaneous transplantation (Fig. 3f and Supplementary Fig. 7). Re-expression of an shRNA-insensitive *Nkx2-1* cDNA (*Nkx2-1*^{*}) reverted the phenotypic alterations elicited by *shNkx2-1*, confirming that the effects of *shNkx2-1* were specifically due to *Nkx2-1* knockdown (Supplementary Fig. 7). Finally, we induced tumours in *Kras*^{LSL-G12D/+}; *p53*^{fllox/fllox} mice with either Lenti-Cre or a lentiviral vector expressing both *Nkx2-1* and Cre (Lenti-*Nkx2-1*/Cre). Expression of exogenous *Nkx2-1* limited tumour progression, resulting in fewer tumours of advanced histopathological grades (Fig. 3g).

To discover *Nkx2-1*-regulated genes, we compared gene expression in T_{nonMet} and T_{nonMet} -*shNkx2-1* cells. Overlapping this gene list with the genes expressed at different levels in T_{nonMet} versus T_{Met} /Met cells uncovered high priority candidate genes (Supplementary Fig. 8). We elected to focus on *Hmga2* given its role in altering global gene expression through the regulation of chromatin structure and its association with embryonic and adult stem-cell states^{21–24} as well as malignant tumours of diverse origins^{25–29}. *Hmga2* was de-repressed by *Nkx2-1* knockdown in T_{nonMet} cells, and regions of *Kras*^{G12D/+}; *p53*^{d/d} tumours that lacked *Nkx2-1* expression were almost universally *Hmga2*⁺ (Fig. 4a–c). Importantly, *Nkx2-1*[−] areas of known metastatic primary tumours and metastases were also *Hmga2*⁺ (Supplementary Fig. 9 and data not shown). Additionally, *Hmga2* was downregulated in T_{Met} cells after expression of *Nkx2-1* cDNA and in T_{nonMet} -*shNkx2-1* cells after re-expression of *Nkx2-1*^{*} (data not shown).

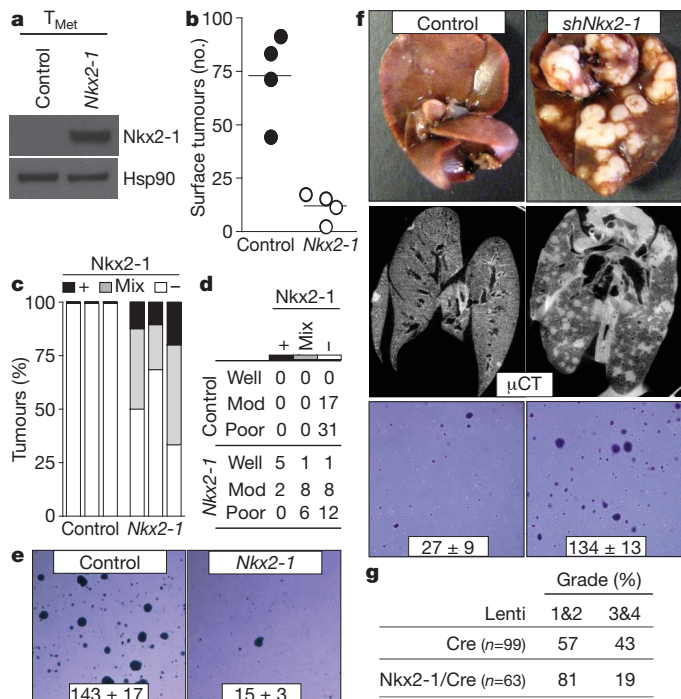


Figure 3 | Nkx2-1 controls lung adenocarcinoma differentiation and restricts metastatic ability. **a**, Exogenous Nkx2-1 protein expression in T_{Met} cells. **b**, Nkx2-1 expression reduces lung nodule formation after intravenous transplantation; $P < 0.002$. **c**, Quantification of Nkx2-1 in lung nodules after T_{Met} or T_{Met}-Nkx2-1 transplantation; $n = 3$ per group. **d**, Association of Nkx2-1 expression with differentiation state after T_{Met} or T_{Met}-Nkx2-1 transplantation. Fisher's exact test on the association of differentiation state with Nkx2-1, $P < 0.002$. **e**, Nkx2-1 expression reduces anchorage-independent growth of T_{Met} cells. Representative images and colony number (mean \pm s.d. of quadruplicate wells, $P < 0.0001$) are shown. **f**, Nkx2-1 knockdown increases liver nodules after intrasplenic injection (top) and lung nodules after intravenous transplantation (middle) of T_{nonMet} cells. Results are representative of 7 mice per group. **g**, *shNkx2-1* enhanced anchorage-independent growth of T_{nonMet} cells (bottom). Representative images and colony number (mean \pm s.d. of triplicate wells, $P < 0.0001$) are shown. **g**, Induction of tumours in *Kras*^{G12D/+}; *p53*^{flx/flx} mice with Nkx2-1/Cre lentivirus reduces the development of advanced tumours (grades 3 and 4). Numbers indicate percentage of tumours in each group.

Although *Hmga2* can be regulated by the Let7 family of microRNAs^{21,25,30}, Let7 levels, *Lin28a* expression and Let7 activity were equivalent in T_{nonMet}, T_{Met} and Met cell lines and were unaltered in T_{nonMet}-*shNkx2-1* cells (Supplementary Fig. 10 and data not shown). *Hmga2* promoter activity was de-repressed in T_{nonMet}-*shNkx2-1* cells and repressed in T_{Met}-Nkx2-1 cells, indicating that expression of *Hmga2* in lung adenocarcinoma cells is regulated, at least in part, through differential promoter activity (Supplementary Fig. 10).

We hypothesized that lung adenocarcinomas progress from an Nkx2-1⁺Hmga2⁻ to an Nkx2-1⁻Hmga2⁺ state. However, metastatic and non-metastatic tumours could be fundamentally distinct at the time of initiation. *Hmga2* is highly expressed in embryonic lung but not in any normal adult lung cells, and early after initiation, *Kras*^{G12D/+}; *p53*^{Δ/Δ} tumours were uniformly Nkx2-1⁺Hmga2⁻ (Fig. 4d and Supplementary Fig. 11). *Kras*^{G12D/+}; *p53*-proficient tumours, which maintain their differentiated phenotype and never metastasize even late after tumour initiation⁵, were almost universally Nkx2-1⁺Hmga2⁻ (Supplementary Fig. 11). Poorly differentiated areas of *Kras*^{G12D/+}; *p53*^{Δ/Δ} tumours with reduced Nkx2-1 expression were never found as *in situ* lesions and were almost always associated with lower grade Nkx2-1-expressing areas (Supplementary Fig. 6). Finally, we induced *Kras*^{G12D/+}; *p53*^{Δ/Δ} tumours with a pool of lentiviral vectors that contain nucleotide barcodes. Amplification and sequencing of the lentivirus-encoded barcodes from adjacent low-grade Nkx2-1⁺Hmga2⁻ and high-grade Nkx2-1⁻Hmga2⁺

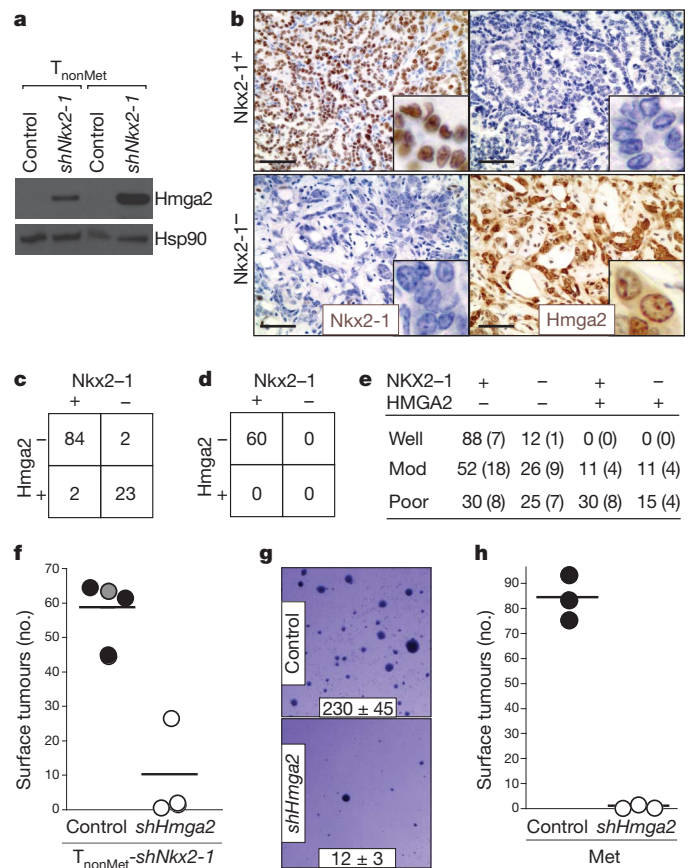


Figure 4 | Nkx2-1 regulates the expression of Hmga2 in advanced lung adenocarcinoma. **a**, Nkx2-1 knockdown de-represses *Hmga2* in T_{nonMet} cell lines. **b**, *Hmga2* and Nkx2-1 are reciprocally expressed in *Kras*^{G12D/+}; *p53*^{Δ/Δ} murine lung adenocarcinomas. Scale bar = 50 μ m. Inset images show cellular features and protein localization. **c**, *Hmga2* and Nkx2-1 expression in advanced *Kras*^{G12D/+}; *p53*^{Δ/Δ} murine lung adenocarcinomas. Fisher's exact test, $P < 10^{-11}$. **d**, Early *Kras*^{G12D/+}; *p53*^{Δ/Δ} tumours are Nkx2-1⁺Hmga2⁻. **e**, NKX2-1 and HMG2 expression in human lung adenocarcinomas. Numbers indicate percentages; numbers in parentheses indicate absolute numbers. **f**, *Hmga2* knockdown reduces the tumorigenic potential of T_{nonMet}-*shNkx2-1* cells after intravenous transplantation. Control samples include the parental T_{nonMet}-*shNkx2-1* cells (grey circle) and cells infected by a control retrovirus (black circles). $P < 0.003$. **g**, *shHmga2* reduces anchorage-independent growth of a metastasis-derived cell line (Met). Representative images and colony number (mean \pm s.d. of quadruplicate wells, $P < 0.0001$) are shown. **h**, *shHmga2* reduces the tumour-seeding potential of a Met cell line after intravenous transplantation; $P < 0.0001$.

areas showed that these areas were clonally related (Supplementary Fig. 12). Although alternative mechanisms leading to the generation of clonally related but phenotypically distinct tumour cell populations are possible, including the expansion of rare Nkx2-1⁻ cells that pre-exist within the tumour, we believe that our data strongly suggest that lung adenocarcinomas undergo a transition from an Nkx2-1⁺Hmga2⁻ state to a more aggressive Nkx2-1⁻Hmga2⁺ state. Our data additionally indicate that an Nkx2-1-dependent gene expression program is a key regulator of this transition.

We next analysed the expression of NKX2-1 and HMGA2 in human adenocarcinoma. Although the expression patterns were diverse, two important conclusions could be made. First, tumours of NKX2-1⁺HMGA2⁻ and NKX2-1⁻HMGA2⁺ phenotypes exist within the spectrum of human lung adenocarcinoma (Fig. 4e and Supplementary Fig. 10). Second, there was a trend towards well-differentiated tumours being NKX2-1⁺HMGA2⁻ whereas moderately and poorly differentiated tumours more often exhibited other combinations of these proteins. Most notably, the moderately and poorly differentiated groups contained NKX2-1⁻HMGA2⁺ tumours (Fig. 4e). These results

underscore the diversity within this single human tumour type and indicate that our genetically defined model probably represents, at the molecular level, a subset of these tumours.

Next we knocked down Hmga2 in T_{nonMet}-shNkx2-1 cells and found that their metastasis seeding potential was greatly reduced after transplantation (Fig. 4f and Supplementary Fig. 13). Additionally, Hmga2 knockdown in a metastasis-derived cell line reduced its anchorage-independent growth and tumour seeding ability after transplantation (Fig. 4g, h and Supplementary Fig. 13). A future challenge will be to understand the molecular mechanism by which Hmga2 controls lung adenocarcinoma metastatic potential. The expansion of Nkx2-1⁺Hmga2⁺ regions within primary lung tumours indicates the acquisition of phenotypes that are advantageous to the primary tumours and also increase the probability of metastatic spread.

NKX2-1 can have both oncogenic and tumour suppressive functions in lung cancer, presumably illustrating context-dependent functions within individual tumours of the same type. Lung adenocarcinomas may differ in their cell of origin, mutation spectrum, or gene expression, leading to distinct requirements for continued NKX2-1 expression and different capacity to tolerate or benefit from NKX2-1 downregulation. Our studies uncovered the molecular and cellular basis for the association of NKX2-1 expression with good patient outcome^{17,18} and HMGA2 expression with poor patient outcome^{26,27}. Our results emphasize the power of genetically engineered mouse models of advanced disease, used in conjunction with human studies, to elucidate mechanisms that control cancer progression and metastatic spread. Through this approach we identified one molecular mechanism by which a highly prevalent tumour type can progress to its malignant state.

METHODS SUMMARY

Mice, tumour initiation and derivation of cell lines. *Kras*^{LSL-G12D}, *p53*^{fllox}, *p53*^{LSL-R270H} and *p53*^{LSL-R172H} mice have been described^{3,5,19}. Tumours were initiated by intratracheal infection of mice with a lentiviral vector expressing Cre recombinase¹⁰. The MIT Institutional Animal Care and Use Committee approved all animal studies and procedures. Cell lines were created by enzymatic and mechanical dissociation of individual lung tumours and metastases harvested from mice 8–14 months after tumour initiation.

LM-PCR, Southern blotting and gene expression analysis. LM-PCR was performed with forward primers specific for the lentiviral LTR. Southern blotting used a Cre probe and standard methods. RNA was extracted using Trizol, analysed for RNA integrity, and prepared with Affymetrix GeneChip WT Sense Target Labelling and Control Reagents kits, followed by hybridization to Affymetrix GeneChip Mouse Exon 1.0 ST Arrays.

Protein and RNA analysis. Western blotting used standard methods and antibodies to Nkx2-1 (Epitomics, Inc.), Hmga2 (BioCheck, Inc.) and Hsp90 (BD Transduction Laboratories) as a loading control. Immunohistochemistry was performed on formalin-fixed, paraffin-embedded 4-μm sections using the ABC Vectastain kit (Vector Laboratories) with antibodies described above. Sections were developed with DAB and counterstained with haematoxylin.

Gene expression and knockdown. Nkx2-1 was stably knocked down with a pLKO-based lentiviral vector (OpenBiosystems/TRC). MSCV-Puro retroviral vectors were used for stable expression of Nkx2-1 and Nkx2-1* (created with four silent mutations using QuikChange Lightning Site-Directed Mutagenesis (Stratagene)). Hmga2 was stably knocked down with an MSCV-Hygro retroviral vector.

Transplantation experiments. For intravenous transplantation, 10⁵ cells resuspended in 200 μl PBS were injected into the lateral tail vein. For intrasplenic transplantation 10⁵ cells resuspended in 50 μl PBS were injected. In all graphs each circle represents an individual mouse and the bar represents the mean. Statistical significance was determined using the Student's *t*-test.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 June 2010; accepted 31 January 2011.

Published online 6 April 2011.

1. Rodenhuis, S. *et al.* Incidence and possible clinical significance of K-ras oncogene activation in adenocarcinoma of the human lung. *Cancer Res.* **48**, 5738–5741 (1988).
2. Takahashi, T. *et al.* p53: a frequent target for genetic abnormalities in lung cancer. *Science* **246**, 491–494 (1989).
3. Jonkers, J. *et al.* Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. *Nature Genet.* **29**, 418–425 (2001).

4. Jackson, E. L. *et al.* The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res.* **65**, 10280–10288 (2005).
5. Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.* **15**, 3243–3248 (2001).
6. Weir, B. A. *et al.* Characterizing the cancer genome in lung adenocarcinoma. *Nature* **450**, 893–898 (2007).
7. Tanaka, H. *et al.* Lineage-specific dependency of lung adenocarcinomas on the lung developmental regulator TTF-1. *Cancer Res.* **67**, 6007–6011 (2007).
8. Kendall, J. *et al.* Oncogenic cooperation and complication of developmental transcription factor genes in lung cancer. *Proc. Natl Acad. Sci. USA* **104**, 16663–16668 (2007).
9. Kwei, K. A. *et al.* Genomic profiling identifies TTF1 as a lineage-specific oncogene amplified in lung cancer. *Oncogene* **27**, 3635–3640 (2008).
10. DuPage, M., Dooley, A. L. & Jacks, T. Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nature Protocols* **4**, 1064–1072 (2009).
11. Monti, S., Tamayo, P., Mesirov, J. & Golub, T. Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Mach. Learn.* **52**, 91–118 (2003).
12. Shedden, K. *et al.* Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nature Med.* **14**, 822–827 (2008).
13. Nguyen, D. X. *et al.* WNT/TCF signaling through LEF1 and HOXB9 mediates lung adenocarcinoma metastasis. *Cell* **138**, 51–62 (2009).
14. Kimura, S. *et al.* The T/ebp null mouse: thyroid-specific enhancer-binding protein is essential for the organogenesis of the thyroid, lung, ventral forebrain, and pituitary. *Genes Dev.* **10**, 60–69 (1996).
15. Krude, H. *et al.* Choroathetosis, hypothyroidism, and pulmonary alterations due to human NKX2-1 haploinsufficiency. *J. Clin. Invest.* **109**, 475–480 (2002).
16. Maeda, Y., Dave, V. & Whitsett, J. A. Transcriptional control of lung morphogenesis. *Physiol. Rev.* **87**, 219–244 (2007).
17. Barletta, J. A. *et al.* Clinical significance of TTF-1 protein expression and TTF-1 gene amplification in lung adenocarcinoma. *J. Cell Mol. Med.* **13**, 1977–1986 (2009).
18. Berghmans, T. *et al.* Thyroid transcription factor 1—a new prognostic factor in lung cancer: a meta-analysis. *Ann. Oncol.* **17**, 1673–1676 (2006).
19. Olive, K. P. *et al.* Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome. *Cell* **119**, 847–860 (2004).
20. Ben-Porath, I. *et al.* An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nature Genet.* **40**, 499–507 (2008).
21. Nishino, J., Kim, I., Chada, K. & Morrison, S. J. Hmga2 promotes neural stem cell self-renewal in young but not old mice by reducing p16^{Ink4a} and p19^{Arf} expression. *Cell* **135**, 227–239 (2008).
22. Yu, F. *et al.* let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell* **131**, 1109–1123 (2007).
23. Li, O., Vasudevan, D., Davey, C. A. & Droge, P. High-level expression of DNA architectural factor HMGA2 and its association with nucleosomes in human embryonic stem cells. *Genesis* **44**, 523–529 (2006).
24. Rommel, B. *et al.* HMGI-C, a member of the high mobility group family of proteins, is expressed in hematopoietic stem cells and in leukemic cells. *Leuk. Lymphoma* **26**, 603–607 (1997).
25. Fusco, A. & Fedele, M. Roles of HMGA proteins in cancer. *Nature Rev. Cancer* **7**, 899–910 (2007).
26. Meyer, B. *et al.* HMGA2 overexpression in non-small cell lung cancer. *Mol. Carcinog.* **46**, 503–511 (2007).
27. Hristov, A. C. *et al.* HMGA2 protein expression correlates with lymph node metastasis and increased tumor grade in pancreatic ductal adenocarcinoma. *Mod. Pathol.* **22**, 43–49 (2009).
28. Rogalla, P. *et al.* Expression of HMGI-C, a member of the high mobility group protein family, in a subset of breast cancers: relationship to histologic grade. *Mol. Carcinog.* **19**, 153–156 (1997).
29. Di Cello, F. *et al.* HMGA2 participates in transformation in human lung cancer. *Mol. Cancer Res.* **6**, 743–750 (2008).
30. Lee, Y. S. & Dutta, A. The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. *Genes Dev.* **21**, 1025–1030 (2007).

Acknowledgements We thank A. Dooley, N. Dimitrova, T. Oliver and M. Ebert for providing reagents; M. Luo (Biology/Koch Institute BioMicro Core) for array processing; M. Kumar for experimental assistance; T. Staton, D. McFadden, A. Shaw and the entire T.J. laboratory for comments. M.M.W. was a Merck Fellow of the Damon Runyon Cancer Research Foundation and was funded by a Genentech Postdoctoral Fellowship. R.G.W.V. is supported by a Fellowship from the Dutch Cancer Society KWF. E.L.S. is supported by a training grant (T32-HL007627). D.Y.C. is supported by an Alfred P. Sloan Foundation Research Fellowship. D.M.F. is a Leukemia and Lymphoma Postdoctoral Fellow. This work was supported by National Institutes of Health grants U01-CA84306 (to T.J.) and K99-CA151968 (to M.M.W.), the Howard Hughes Medical Institute, the Ludwig Center for Molecular Oncology at MIT, and in part by the Cancer Center Support (core) grant P30-CA14051 from the National Cancer Institute. T.J. is the David H. Koch Professor of Biology and a Daniel K. Ludwig Scholar.

Author Contributions M.M.W. and T.J. designed the study; M.M.W., T.L.D., C.K.-K. and E.L.S. performed experiments; R.G.W.V., M.M.W., C.A.W., D.D.H., S.H. and D.Y.C. conducted bioinformatic analyses; E.L.S. and R.T.B. provided pathology assistance; S.Y. and D.C. provided technical assistance; M.J.D. provided reagents; D.M.F. and M.M. gave conceptual advice; M.M.W. and T.J. wrote the paper with comments from all authors.

Author Information Gene expression data was deposited in Gene Expression Omnibus (GSE26874). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.J. (tjacks@mit.edu).

METHODS

Mice, tumour initiation and creation of cell lines. *Kras*^{LSL-G12D}, *p53*^{fllox}, *p53*^{LSL-R270H} and *p53*^{LSL-R172H} mice have been described^{33–5,19,31}. Tumours were initiated by intratracheal infection of mice with a lentiviral vector expressing Cre recombinase¹⁰. The lenti-Cre vector was co-transfected with packaging vectors (delta8.2 and VSV-G; gifts from D. Trono³²) into 293T cells using TransIT-LT1 (Mirus Bio). The resultant supernatant was collected at 48 and 72 h. Virus was recovered by ultracentrifugation at 25,000 r.p.m. for 90 min and re-suspended in an appropriate volume of PBS. The MIT Institutional Animal Care and Use Committee approved all animal studies and procedures. Cell lines were created from individual lung tumours and metastases harvested from mice 8–14 months after tumour initiation. Tumours were cut into small pieces and then digested for 30 min at 37 °C in 2 ml of HBSS-free containing trypsin, collagenase IV and dispase in a 15-ml conical tube. Following digestion, 4 ml of Quench Solution (L15 media with 400 µl of FBS and 15 µl of 5 mg ml⁻¹ DNase) was added. Digested tumour samples were then pressed through 40 µm cell strainers (BD Biosciences). Finally, samples were centrifuged at 1,000 r.p.m. for 5 min, re-suspended in culture media (DMEM, 10% FBS, penicillin/streptomycin, glutamine), and plated in a 12-well plate. Cells were washed and culture media was changed every day for a week until stable cell lines were formed. In general we were able to derive cell lines from approximately 50% of tumours and metastases without a noticeable difference in the frequency of cell-line generation from metastases from different sites relative to the primary lung tumours.

Linker-mediated PCR. Linker-mediated PCR (LM-PCR) was performed essentially as previously described³³. Annealing two oligonucleotides 5'-TAGTCCC TTAAGCGGAG-3' and 5'-GTAATACGACTCACTATAGGGCTCCGCTTAA GGGAC-3' created the linker. This linker was ligated to MseI-digested genomic DNA followed by nested PCR using the following primers: forward 1 (lentivirus) 5'-CTCAATAAAGCTTGCCTTG-3', reverse 1 (linker) 5'-GTAATACGACTC ACGATAGGGC-3'; forward 2 (lentivirus) 5'-CTGTTGTGTGACTCTGGT AAC-3', reverse 2 (linker) 5'-AGGGCTCCGCTTAAGGGAC-3'.

The PCR products were separated on an agarose gel and specifically amplified bands were gel extracted, cloned into pCRII-TOPO (Invitrogen), and sequenced to identify the exact genomic locus of lentiviral integration. Specific primers were designed in that genomic region to amplify a product from the lentiviral genome to the adjacent somatic genome (lenti-genome band) and also to amplify a product from the other non-integrated allele of that genomic locus (genome-genome band) (see Fig. 1b). Thus, related tumours can be identified by this three-primer PCR reaction with related tumours having both the lenti-genome and genome-genome bands and unrelated tumours having just the genome-genome band. In general we detected a single lentiviral integrated in each tumour. When we have cloned the exact position of the lentiviral insertion sites we have not yet found an insertion site near a gene of interest. Therefore, it is extremely unlikely that these lentiviruses function as insertional mutagens.

Southern blotting. For Southern blotting, 30 µg of genomic DNA was doubly digested with EcoRI and BamHI overnight at 37 °C. 1 µl of enzyme was added in the morning and DNA was digested for an additional hour. The concentration of the digested DNA was determined using a NanoDrop spectrometer. The digested DNA was loaded onto a 0.7% agarose gel and samples were electrophoresed at 30 V for 17–20 h. DNA was transferred onto a positively charged nylon membrane (Hybond-XL, Amersham) overnight and then ultraviolet-crosslinked. A probe for Cre was obtained by PCR amplification from a Cre-containing plasmid using the following primers: forward primer 5'-GCTCTAGCGTTCGAACGCAC-3', reverse primer 5'-GCTGGCCGGCCCATCGCCATCTTCCAGCAGGC-3'.

Probes were labelled using a DECAprime II Random Priming DNA labelling kit, according to the manufacturer's instructions (Ambion). The membrane was hybridized with labelled probe at 65 °C overnight, then washed and exposed overnight on a PhosphorImager Screen, and imaged using a STORM 860 Molecular Imager.

Murine gene expression profiling and analysis. Twenty-three samples from murine primary tumour- and metastasis-derived cell lines were collected and RNA was isolated. RNA was extracted using Trizol (Invitrogen), analysed for RNA integrity, and prepared with Affymetrix GeneChip WT Sense Target Labelling and Control Reagents kit, followed by hybridization to Affymetrix GeneChip Mouse Exon 1.0 ST arrays. The resulting image files were pre-processed using the aroma.affymetrix³⁴ and FIRMA libraries³⁵ available in the R/Bioconductor software environment^{36,37}. Probe intensities were summarized as expression levels using quantile normalization and RMA.

We applied hierarchical consensus clustering with complete linkage¹¹ and 1 – (Pearson's correlation coefficient) as a distance metric, to identify consistently similar expression profiles. The procedure was run over 2,000 iterations with a subsampling ratio of 0.8 on all 23 murine lung tumour samples, and 1,500 variably expressed genes as determined by the median absolute deviation. To identify

differentially expressed genes between the T_{nonMet} and $T_{\text{Met}}/\text{Met}$ samples the significance analysis of microarrays algorithm³⁸ was applied.

To determine gene expression alterations induced by Nkx2-1 knockdown in T_{nonMet} cells, Affymetrix GeneChip Mouse Exon 1.0 ST Array analyses were performed on 368T1-Control, 368T1-*shNkx2-1*, 394T4-Control and 394T4-*shNkx2-1* cells. The 394T4 samples were run in duplicate using either the Affymetrix GeneChip WT Sense Target Labelling or the Nugene Ovation Systems Target Preparation kits. Thus, a pairwise comparison between the three control and their corresponding *shNkx2-1* data sets was used to determine the potential Nkx2-1-regulated genes (Supplementary Table 2). We overlapped the gene list generated by the control versus *shNkx2-1* comparisons ($\log_2 > 0.4$, paired *t*-test < 0.08) with the gene list generated from the T_{nonMet} versus $T_{\text{Met}}/\text{Met}$ samples ($\log_2 > 0.9$, unpaired *t*-test < 0.02) to identify high priority targets.

Two curated gene sets of proliferation-related genes in humans were obtained from the Molecular Signatures Database (<http://www.broadinstitute.org/gsea/msigdb/index.jsp>, cell_proliferation (232 genes, renamed PROLIFERATION_GENES1) and proliferation_genes (394 genes, renamed PROLIFERATION_GENES2)). The majority of these human genes mapped to mouse genes, resulting in two proliferation signatures for mice, with 187 and 332 genes respectively. These two signatures were projected on the $T_{\text{Met}}/\text{Met}$ and T_{nonMet} samples and Nkx2-1 knockdown cell lines data sets. No significant difference in signature scores was observed in T_{nonMet} versus $T_{\text{Met}}/\text{Met}$ or T_{nonMet} -Control versus $T_{\text{nonMet}}\text{-shNkx2-1}$ comparisons as determined by a Student's *t*-test.

Pre-processing of human data set and projection of gene signatures. Human lung adenocarcinoma gene expression analysis used raw data from ref. 12 (364 patients) and ref. 13 (129 patients). These data sets contain patients of all stages. Probes intensities from the Affymetrix U133A platform used in these studies were analysed in a gene-centric fashion³⁹ and preprocessed using quantile normalization and RMA.

To verify findings from our mouse model, the T_{nonMet} signature and a signature from human embryonic stem cells^{20,40} were projected on the expression profiles of lung adenocarcinoma samples from 364 patients¹² and 129 patients¹³. For a given human adenocarcinoma sample, gene expression values were rank-normalized and rank-ordered. The Empirical Cumulative Distribution Functions (ECDF) of the genes in the gene signature and the remaining genes were calculated. By an integration of the difference between the ECDFs, a statistic was calculated similar to the one used in Gene Set Enrichment Analysis but based on absolute expression rather than differential expression. Details on the method can be found in two papers in which it was applied^{41,42}. The statistic is calculated by replacing the absolute gene expression levels for a given signature *G* in a single sample *S*_{*i*} by their ranks $L = \{r_1, r_2, r_3, \dots, r_N\}$ and rank ordering them. By taking the weighted sum of the difference between the ECDF of the genes in the signature *P_G* and the ECDF of the remaining genes *P_{NG}* an enrichment score ES(*G*, *S*) is obtained:

$$ES(G, S) = \sum_i [P_G(G, S, i) - P_{NG}(G, S, i)]$$

$$P_G(G, S, i) = \sum_{r_j \in G \& j \leq i} \frac{|r_j|^{1/4}}{\sum_{r_j \in G} |r_j|^{1/4}}, P_{NG}(G, S, i) = \sum_{r_j \notin G \& j \leq i} \frac{1}{(N - N_G)}$$

This calculation was repeated for the T_{nonMet} and ES1 signatures and each sample in the data set. The exponent $1/4$ adds a slight weight proportional to the outcome value, which is signed.

Human lung adenocarcinoma tissue arrays. To determine the expression of NKX2-1 and HMGA2 in human lung adenocarcinoma we stained a tissue microarray (US BioMax, LC1921) for these proteins using the antibodies described above. Differentiation state of the tumours was provided by US BioMax and confirmed by a pathologist (E.L.S.).

qPCR. Quantitative RT-PCR was performed on Trizol-extracted RNA using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems). qPCR reactions were performed using SYBR Green Jumpstart Taq Ready Mix (Sigma) and a ABI Prism thermocycler (Applied Biosystem). *Nkx2-1* expression is shown relative to *Gapdh* control. qPCR primers were: *Nkx2-1* forward 5'-AAACTGCG GGGATCTGAG-3', *Nkx2-1* reverse 5'-TGCTTTGGACTCATCGACAT-3'; *Gapdh* forward 5'-TTTGATGTTAGTGGGTCTCG-3', *Gapdh* reverse 5'-AGC TTGTCATCAACGGGAAG-3'.

Immunohistochemical analysis and quantification. Samples for histology were fixed in 3.7% formalin in PBS for 24 h and stored in 70% ethanol until paraffin embedding. Immunohistochemistry was performed on formalin-fixed, paraffin-embedded 4-µm sections using the ABC Vectastain kit (Vector Laboratories) with antibodies to Nkx2-1 (Epitomics), Hmga2 (BioCheck), p63 (Lab Vision), cytokeratin 8 (Developmental Studies Hybridoma Bank), and phospho-histone3 (Cell

Signaling). Sections were developed with DAB and counterstained with haematoxylin. Tumour size (mm²) and mitotic index (number of phospho-H3 cells/mm²) after transplantation of T_{Met} and T_{Met-Nkx2-1} cells were quantified using BioQuant Software. Haematoxylin and eosin staining was performed using standard methods.

cDNA expression and knockdown. pLKO.shNkx2-1: to knock down *Nkx2-1*, shRNA pLKO.1 lentiviral vectors targeting *Nkx2-1* were purchased from OpenBiosystems and are part of The RNAi Consortium (TRC)⁴³. The best hairpin sequence targeting *Nkx2-1* was *shNkx2-1* (TRCN0000086266) 5'-CGCCATGTCTGTCTACCTT-3'. An shRNA targeting luciferase was used as a control. Lentiviral production was performed as described above⁴⁴. The pLKO.1 vectors were co-transfected with packaging vectors (delta8.2 and VSV-G, gifts from D. Trono³²) into 293T cells as described above. The resultant supernatant was collected at 48 and 72 h. For cell line infection, 3 × 10⁵ cells were plated in a well of a 6-well plate and incubated at 37 °C for ≥16 h. Cells were incubated with equal volumes of virus and complete media for at least 24 h, at which point fresh media was added. Puromycin (16 µg ml⁻¹ for 368T1, 32 µg ml⁻¹ for 394T4) was also added at this point. Cells were collected for isolation of RNA and protein 2 weeks after infection with the shRNA viruses.

MSCV-driven ecotropic retroviral vectors were made by transfecting 293T cells with equal amount of the MSCV vector and pCLEco package vector⁴⁵.

MSCV-Puro (control): retroviral expression vector driving the cDNA of interest off the MSCV LTR and puromycin resistance off the PGK-promoter (Clontech). The empty vector was used as a control.

MSCV-Nkx2-1/Puro: *Nkx2-1* cDNA was PCR amplified using the following primers: forward 5'-GAGTTAACCACCATGTCGATGAGTCCAAAGCAC-3', reverse 5'-CTGAATTCTCACCAGGTCCGACCATAAAG-3', followed by cloning into pCRII-TOPO (Invitrogen), sequence verification, digestion and ligation into MSCV-Puro.

MSCV-Nkx2-1*/Puro: an *shNkx2-1* insensitive *Nkx2-1* cDNA was created by engineering four silent mutations using QuikChange Lightning Site-Directed Mutagenesis (Stratagene) and the following primers: 5'-GGTCCGACCATAAAGCAAAGTGGAGCAGGACATGGCGCCATAGTCCGAG-3' and 5'-CTCGGACTATGGCGCCATGTCCTGCTCCACTTTGCTTTATGGTCGGACC-3' followed by sequence verification and cloning into the MSCV-Puro retroviral expression vector.

MSCV-Hygro (control): a hygromycin-resistant retroviral expression vector was created by swapping the hygromycin-resistance gene for the puromycin-resistance gene in MSCV/LTRmiR30-Puro-IRES-GFP⁴⁶. The empty vector was used as a control.

MSCV-shHmga2/Hygro: shRNAs were designed using <http://www.biopredsi.org/start.html> and resources available through G. Hannon's laboratory at Cold Spring Harbour Laboratories (<http://katahdin.cshl.org/siRNA/RNAi.cgi?type=shRNA>). Five sequences within *Hmga2* were cloned into the miR30 sequence and tested for effective knockdown. The oligonucleotides were amplified, digested and cloned into MSCV-Hygro. The correct sequence was verified by sequencing the final vector. The following oligonucleotide gave the best *Hmga2* knockdown: *shHmga2* (NM_010441-1375) 5'-TGCTGTTGACAGTGAGCGAAAGGACTATATTAATCACCTTTAGTGAAGCCACAGATGTAAAGTGATTAATATAGTCCTTCTGCCTACTGCCTCGGA-3'. This was used to knock down *Hmga2* in the T_{nonMet}-*shNkx2-1* cells.

pLKO.shHmga2: to knock down *Hmga2* in a metastasis-derived cell line (482N1) pLKO.1 lentiviral vectors targeting *Hmga2* were obtained from The RNAi Consortium⁴³. The best hairpin sequence targeting *Hmga2* was *shHmga2* (TRCN0000265760) 5'-GAACTTATCAAGACGATTAA-3'. An shRNA targeting luciferase was used as a control. Lentiviral production was performed as described above⁴⁴.

Cell culture and transplantation assays. To assess proliferation, 100,000 cells were plated in each well of a 6-well plate. Eighteen hours later the sub-confluent cells were labelled with 10 µM BrdU for 1 h followed by anti-BrdU staining using the BD APC flow kit following the manufacturer's instructions. For anchorage-independent growth assays, cells were plated in triplicate or quadruplicate in 35 mm tissue culture dishes in 0.4% agar in culture media on top of a layer of 0.8% agar with culture media. For T_{nonMet} and T_{nonMet}-*shNkx2-1* 3 × 10⁴ cells were plated. For T_{Met}, T_{Met-Nkx2-1}, Met and Met-*shHmga2* 1 × 10⁴ cells were plated. Cells were allowed to grow at 37 °C for 2–3 weeks. Colonies were stained with 0.2% crystal violet at room temperature for 30 min and subsequently destained with water for several days. Once the colonies were visible by eye, they were counted using a microscope. A colony was defined as anything containing more than 10 cells.

For intravenous injection, recipient mice were injected with 10⁵ cells re-suspended in 200 µl PBS in the lateral tail vein. Intravenously injected mice were analysed 3–4 weeks after injection. Quantification of lung tumour nodules was performed by counting all the visible surface tumours on the large left lung lobe.

Intrasplenic injection of 10⁵ cells re-suspended in 50 µl HBSS was performed as described^{47,48}. Briefly, mice were anaesthetized with avertin (0.5 mg g⁻¹ intraperitoneally) before surgery. Once the animals were under deep anaesthesia the fur was removed from the left abdominal and thoracic areas using surgical clippers. The area was disinfected with Betadine and 70% ethanol. The spleen was exposed through a small incision. Cells were injected into the spleen with a single injection using an insulin syringe. Cells were given 10 min to travel through the vasculature to the liver, after which the entire spleen was removed to prevent the formation of a large splenic tumour mass. To remove the spleen, a dissolvable 4-0 suture was tied snugly around the base of the spleen including the major splenic vasculature and the spleen was removed. The muscle wall was closed with 4-0 dissolvable sutures and the skin incision closed with sterile 7-mm wound clips (Roboz). Intraperitoneally injected mice were analysed 3–4 weeks after injection. Quantification of liver tumour nodules was performed by counting all the visible surface tumours under a dissecting scope.

For T_{Met} and T_{Met-Nkx2-1} subcutaneous injection, recipient mice were injected with 10⁵ cells re-suspended in 50 µl PBS under the skin on their hind flank. Subcutaneously injected mice were analysed 10 weeks after injection. Both the presence of a tumour nodule and the weight of the resulting tumour were scored. For T_{nonMet} and T_{nonMet}-*shNkx2-1* subcutaneous injections, recipient mice were injected with 10⁶, 10⁵, 10⁴, or 10³ cells re-suspended in 50 µl PBS under the skin on each of their hind flanks. Subcutaneously injected mice were analysed 4 weeks after injection. Both the presence of a tumour nodule and the weight of the resulting tumour were scored.

For all transplants recipient mice were 129/B16 F₁ mice (Jackson Laboratories) except in experiments with MSCV-*shHmga2*/Hygro and MSCV-Hygro (control), which necessitated the use of immunocompromised *Rag2*^{-/-} recipient mice, due to an apparent immunogenicity of the hygromycin-resistant gene.

Creation of dual promoter lentiviral vector for *Nkx2-1* cDNA expression. To determine the effect of continued exogenous *Nkx2-1* expression on lung tumour progression we created a lentiviral vector that expressed both *Nkx2-1* and Cre-recombinase. This vector had the *Ubc* promoter driving *Nkx2-1* cDNA and the PGK promoter driving Cre expression (pLL3.0-Ubc-Nkx2-1;Pgk-Cre). A lentivirus expressing Cre alone (PGK-Cre) was used as a control. Virus was produced and titred as described above. Cohorts of *Kras*^{LSL-G12D}; *p53*^{flox/flox} mice were infected with each lentiviral vector. Tumour grade was determined 18 weeks after tumour initiation.

DNA copy number data set. DNA copy number analysis was performed on 25 lung adenocarcinoma cell lines using the Illumina Genomic DNA Sample Preparation Kit (unpublished data set). Single-end 35 nucleotide reads were generated using the Illumina Genome Analyser IIX. Reads were aligned to the July 2007 build of the *Mus musculus* genome (NCBI37/mm9) with MAQ and filtered for mapping quality >30. Copy number ratios were calculated as the number of normalized reads from the tumour sample, divided by the number of normalized reads from the reference129/SVJ strain. Boundaries of copy number changes were identified using change point analysis⁴⁹. DNA copy number was visualized using the Broad Institute's Integrated Genome Viewer (<http://www.broadinstitute.org/igv>).

Hmga2 promoter assays. To assess the activity of a conserved promoter of the *Hmga2* gene, a 2.83-kb fragment of genomic DNA (NCBI37/mm9:chr10:119913940-119916769) was PCR amplified and ligated into the multiple cloning site of the pGL3 firefly luciferase vector (Promega). Tumour cells were transfected with the *Hmga2*-luc construct and a thymidine-kinase promoter driven Renilla luciferase construct using Attractene (Qiagen). Luciferase activity was determined using the Dual Luciferase Reagents (Promega) and *Hmga2* activity was normalized to the Renilla luciferase activity. An empty pGL3 construct was used as a control and the *Hmga2*-luciferase values were additionally normalized to the values obtained using the empty pGL3 construct.

1. Tuveson, D. A. *et al.* Endogenous oncogenic K-ras(G12D) stimulates proliferation and widespread neoplastic and developmental defects. *Cancer Cell* **5**, 375–387 (2004).
2. Trono, D. Lentiviral vectors: turning a deadly foe into a therapeutic agent. *Gene Ther.* **7**, 20–23 (2000).
3. Wu, X., Li, Y., Crise, B. & Burgess, S. M. Transcription start regions in the human genome are favored targets for MLV integration. *Science* **300**, 1749–1751 (2003).
4. Bengtsson, H., Ray, A., Spellman, P. & Speed, T. P. A single-sample method for normalizing and combining full-resolution copy numbers from multiple platforms, labs and analysis methods. *Bioinformatics* **25**, 861–867 (2009).
5. Purdom, E. *et al.* FIRMA: a method for detection of alternative splicing from exon array data. *Bioinformatics* **24**, 1707–1714 (2008).
6. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
7. R Development Core Team. R: A language and environment for statistical computing, reference index version 2.6.1. (<http://www.R-project.org/>) (R Foundation for Statistical Computing, 2005).

38. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).
39. Liu, H. *et al.* AffyProbeMiner: a web resource for computing or retrieving accurately redefined Affymetrix probe sets. *Bioinformatics* **23**, 2385–2390 (2007).
40. Assou, S. *et al.* A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. *Stem Cells* **25**, 961–973 (2007).
41. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
42. Verhaak, R. G. *et al.* Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–110 (2010).
43. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283–1298 (2006).
44. Kumar, M. S. *et al.* Suppression of non-small cell lung tumor development by the let-7 microRNA family. *Proc. Natl Acad. Sci. USA* **105**, 3903–3908 (2008).
45. Naviaux, R. K., Costanzi, E., Haas, M. & Verma, I. M. The pCL vector system: rapid production of helper-free, high-titer, recombinant retroviruses. *J. Virol.* **70**, 5701–5705 (1996).
46. Dickins, R. A. *et al.* Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. *Nature Genet.* **37**, 1289–1295 (2005).
47. Luo, J. L. *et al.* Nuclear cytokine-activated IKK α controls prostate cancer metastasis by repressing Maspin. *Nature* **446**, 690–694 (2007).
48. Morimoto-Tomita, M., Ohashi, Y., Matsubara, A., Tsuiji, M. & Irimura, T. Mouse colon carcinoma cells established for high incidence of experimental hepatic metastasis exhibit accelerated and anchorage-independent growth. *Clin. Exp. Metastasis* **22**, 513–521 (2005).
49. Chiang, D. Y. *et al.* High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nature Methods* **6**, 99–103 (2009).

CPEB and two poly(A) polymerases control miR-122 stability and p53 mRNA translation

David M. Burns¹, Andrea D'Ambrogio¹, Stephanie Nottrott¹ & Joel D. Richter¹

Cytoplasmic polyadenylation-induced translation controls germ cell development^{1,2}, neuronal synaptic plasticity^{3–5} and cellular senescence^{6,7}, a tumour-suppressor mechanism that limits the replicative lifespan of cells^{8,9}. The cytoplasmic polyadenylation element binding protein (CPEB) promotes polyadenylation by nucleating a group of factors including defective in germline development 2 (Gld2), a non-canonical poly(A) polymerase^{10–12}, on specific messenger RNA (mRNA) 3' untranslated regions (UTRs). Because CPEB regulation of *p53* mRNA polyadenylation/translation is necessary for cellular senescence in primary human diploid fibroblasts⁶, we surmised that Gld2 would be the enzyme responsible for poly(A) addition. Here we show that depletion of Gld2 surprisingly promotes rather than inhibits *p53* mRNA polyadenylation/translation, induces premature senescence and enhances the stability of *CPEB* mRNA. The *CPEB* 3' UTR contains two miR-122 binding sites, which when deleted, elevate mRNA translation, as does an antagomir of miR-122. Although miR-122 is thought to be liver specific, it is present in primary fibroblasts and destabilized by Gld2 depletion. Gld4, a second non-canonical poly(A) polymerase, was found to regulate *p53* mRNA polyadenylation/translation in a CPEB-dependent manner. Thus, translational regulation of *p53* mRNA and cellular senescence is coordinated by Gld2/miR-122/CPEB/Gld4.

Mouse embryo fibroblasts (MEFs) derived from *CPEB* knockout mice do not senesce as do MEFs derived from wild-type mice, but instead are immortal. Senescence is rescued when ectopic *CPEB* is expressed in the knockout MEFs and potentiated when expressed in wild-type MEFs⁷. Human foreskin fibroblasts depleted of *CPEB* also bypass senescence and divide for approximately 270 days compared with wild-type cells, which senesce after about 90 days. As with the mouse cells, ectopic expression of *CPEB* rescues senescence in knock-down cells and potentiates senescence in wild-type cells. *CPEB* controls the polyadenylation-induced translation of *p53* mRNA, and indeed *CPEB*-induced senescence requires *p53*. Depletion of *CPEB* also induces the 'Warburg effect', where mitochondrial respiration is reduced and cells produce ATP primarily through glycolysis⁶.

To investigate the possibility that *CPEB* control of *p53* polyadenylation requires Gld2, human primary foreskin fibroblasts were stably transduced with lentiviruses expressing two different short hairpin RNA (shRNAs) against the Gld2 coding sequence. Surprisingly, Gld2 depletion (Fig. 1a, b) induced an increase in both *p53* protein levels (Fig. 1c) and *p53* mRNA polyadenylation (Fig. 1d and Supplementary Fig. 1). Also unexpectedly, depletion of Gld2 resulted in increased oxygen consumption (Fig. 1e) and entry into a senescence-like cell-cycle arrest as evidenced by β -galactosidase staining at acidic pH (Fig. 1f). In comparison, *CPEB*-depleted cells had decreased oxygen consumption, fewer cells staining with β -galactosidase, increased lifespan and, most importantly, reduced poly(A) tail size on *p53* mRNA and approximately 50% reduction in *p53* protein levels⁶.

These paradoxical results prompted us to examine *CPEB* levels in Gld2-deficient cells because *CPEB* is required for normal *p53* mRNA translation⁶. After comparing the amounts of *CPEB* nuclear pre-mRNA

by reverse transcription followed by quantitative PCR (RT-qPCR) and mostly cytoplasmic mRNA by exon-specific RT-qPCR, we found that the pre-mRNA levels, which generally reflect transcription, were nearly unchanged whereas cytoplasmic mRNA levels increased by about five-fold (Fig. 2a). Thus, in the absence of Gld2, *CPEB* mRNA unexpectedly was more stable.

Surmising that Gld2 might control *p53* protein levels through *CPEB*, we next used a *Renilla* luciferase (Rluc) and firefly luciferase (Fluc) reporter system to investigate post-transcriptional regulation of *CPEB* by Gld2. As shown in Fig. 2b, c, the entire *CPEB* 3' UTR was

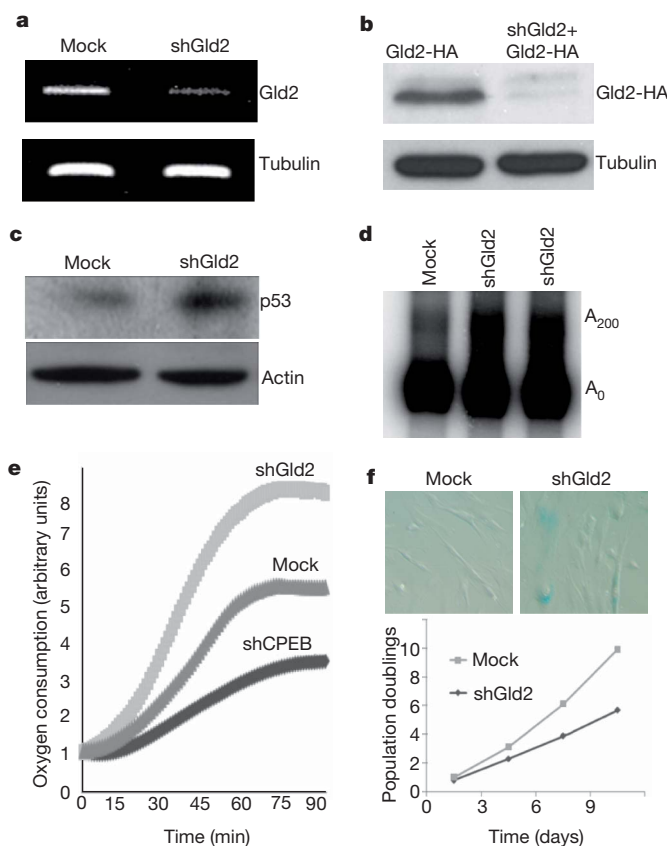


Figure 1 | Depletion of Gld2 enhances *p53* expression. **a**, RT-PCR of Gld2 and tubulin RNAs after infection of human foreskin fibroblasts with lentiviruses expressing shRNA against Gld2 or GFP (Mock, same in all panels). **b**, Knockdown of Gld2-HA in cells expressing ectopic Gld2-HA. Tubulin served as a loading control. **c**, Western blot showing 2.5-fold enhanced expression of *p53* relative to tubulin after Gld2 depletion. **d**, Poly(A) tail analysis of *p53* mRNA in wild-type and Gld2-depleted cells (two shRNAs targeting different regions of Gld2 were used). **e**, Oxygen consumption in cells infected with shCPEB, shGld2, or empty vector (Mock). **f**, Mock or shGld2-infected cells stained for β -galactosidase, which denotes cellular senescence. Population doublings were determined in wild-type or Gld2-depleted cells.

¹Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.

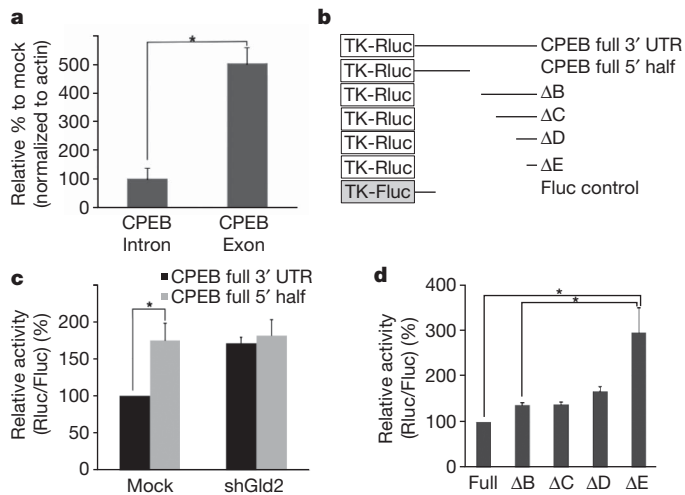


Figure 2 | Gld2 knockdown increases CPEB reporter mRNA and translation by a post-transcriptional mechanism. **a**, Fold change of nuclear (intron-containing) or predominantly cytoplasmic (exon-containing) CPEB RNA after Gld2 depletion ($n = 3$; bars, s.e.m.). **b**, Reporter constructs used in the following experiments (numbers refer to nucleotides of CPEB 3' UTR). **c**, **d**, Cells expressing firefly luciferase (Fluc) as a control and *Renilla* luciferase (Rluc) as noted in **b** were depleted of Gld2; the amount of *Renilla* luciferase activity (relative to firefly) was derived from RNA containing the entire CPEB 3' UTR (full) and set at 100. In all panels, $n = 3$; bars, s.e.m.; * $P < 0.05$, ** $P < 0.01$ (Student's *t*-test).

translated about 40% less efficiently compared with a reporter lacking the 3' most 455 nucleotides (mock). However, in Gld2-deficient cells, the two reporters were translated equally. Additional deletions (Δ) of the CPEB 3' UTR suggested that there might be multiple regions that elicited increases in reporter translation after Gld2 knockdown (that is, ΔE translation was about twofold greater than ΔB , ΔC or ΔD translation) (Fig. 2d).

Analysis of the regions of the CPEB 3' UTR that mediated translational repression by Gld2 revealed the presence of two potential miR-122 binding sites (Supplementary Fig. 2). Although miR-122 is thought to be liver-specific and account for approximately 70% of the total population of microRNAs in that tissue¹³, deletion of these specific sites, either individually or combined, alleviated translational repression in Gld2-depleted cells (Fig. 3a), which were nearly identical to those observed with the large deletions (Fig. 2d). These results suggest that miR-122 might repress CPEB mRNA translation in human skin fibroblasts and indicate that this microRNA (miRNA) is more widely distributed than originally thought. Indeed, recent evidence shows that miR-122 is present in human skin¹⁴ and even HEK293 cells¹⁵.

To assess directly whether miR-122 might repress CPEB mRNA expression, we first cloned and sequenced it from human foreskin fibroblasts and found that it contained a non-templated 3' monoadenylate residue (Fig. 3b; see discussion). Next, cells were electroporated with a locked nucleic acid (LNA) antagomir for miR-122, or as a control, a scrambled LNA. The miR-122 antagomir enhanced reporter expression by about 3.25-fold relative to control (Fig. 3c), but had no stimulatory effect on a reporter whose 3' UTR contained no miR-122 sites (Supplementary Fig. 3). Based on evidence from Katoh *et al.*¹⁶, who demonstrated that, in murine liver, Gld2 is essential for miR-122 stability, we suspected that Gld2 might influence CPEB expression and possibly *p53* mRNA translation by controlling miR-122 steady-state levels. Indeed, Fig. 3d demonstrates that depletion of Gld2 from skin fibroblasts reduced the level of miR-122 by nearly 40-fold. Importantly, when miR-122 LNA antagomir-transduced cells were incubated with the proteasome inhibitor MG132 and pulsed-labelled with [³⁵S]methionine for 15 min followed by *p53* immunoprecipitation, there was a twofold increase in the synthesis rate of *p53* (Fig. 3e, f). Taken together, these data demonstrate that human primary skin

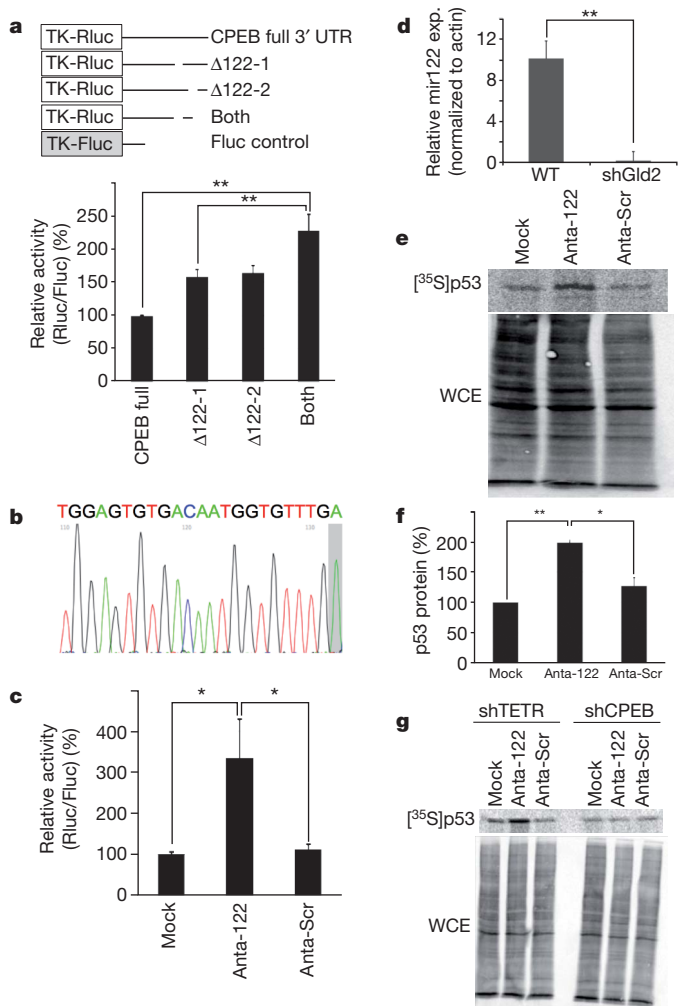


Figure 3 | miR-122 activates *p53* mRNA translation by repressing CPEB. **a**, Gld2-depleted fibroblasts were transduced with firefly and *Renilla* luciferase with CPEB full-length or deletion mutant 3' UTRs lacking putative miR-122 sites (Supplementary Fig. 2). The data are expressed as in Fig. 2; in all panels, $n = 3$; bars, s.e.m.; * $P < 0.05$, ** $P < 0.01$ (Student's *t*-test). **b**, Sequence of miR-122 from fibroblasts; a non-templated adenosine is shaded. **c**, Fibroblasts expressing firefly and *Renilla* luciferase containing the CPEB 3' UTR were electroporated with miR-122 (Anta-122), scrambled (Anta-Scr) or no LNA antagomir (Mock); data are expressed as in Fig. 2. **d**, Quantitative RT-PCR for miR-122 in cells expressing GFP (wild type (WT)) or shGld2. **e**, **f**, Immunoprecipitation of [³⁵S]methionine-labelled *p53* from MG132-treated cells transduced with no (mock-GFP), miR-122 (Anta-122) or scrambled (Anta-Scr) LNA antagomirs. WCE, whole cell lysate. **g**, Fibroblasts were treated as in **e–g** after first expressing either TET repressor (shTETR, a control) or CPEB shRNA.

fibroblasts contain miR-122 and that Gld2 controls its steady state levels or activity.

Although consistent with the hypothesis that miR-122 mediates *p53* mRNA translation through CPEB, these data do not eliminate the possibility that miR-122 could act through another molecule to regulate *p53* synthesis (note that *p53* mRNA has no miR-122 sites according to Targetscan.org or MicroRNA.org). Consequently, we infected cells with a lentivirus expressing shRNA for CPEB as well as the miR-122 antagomir followed by a 15 min pulse of [³⁵S]methionine and *p53* immunoprecipitation. Figure 3g shows that although miR-122 antagomir alone elicited an increase in *p53* synthesis, the antagomir plus shRNA for CPEB induced no increase. Taken together, these data demonstrate that Gld2 activity stabilizes miR-122, which in turn reduces CPEB expression; CPEB then acts directly on *p53* mRNA to control poly(A) tail length and translation.

If not Gld2, what poly(A) polymerase modifies *p53* mRNA polyadenylation and translation? We surmised that an alternative non-canonical poly(A) polymerase, that is, one that lacks an RNA binding domain and thus would require another factor such as CPEB to be tethered to the RNA, would most probably be involved. Two cytoplasmic enzymes have this characteristic: Gld4 (PAPD5)¹⁷ and MitoPAP (PAPD1)¹⁸. Both polymerases were depleted with shRNAs (Supplementary Fig. 4) but only the loss of Gld4 reduced *p53* protein levels (Fig. 4a). To investigate whether Gld4 interacts with *p53* mRNA in a CPEB-dependent manner, Flag-Gld4 was expressed in cells (Supplementary Fig. 5) containing shRNA for tetracycline repressor (TETR, a control) or CPEB. Gld4 was then immunoprecipitated and the extracted RNA was examined for *p53* and GAPDH (a control) RNAs by RT-PCR (Fig. 4b). *p53* mRNA was detected only when CPEB was present, suggesting that Gld4 is anchored to *p53* mRNA by CPEB, and indeed, CPEB co-immunoprecipitated Gld4 but not Mcl1, a non-specific control (Fig. 4c). Finally, depletion of Gld4 reduced *p53* mRNA polyadenylation (Fig. 4d), which probably then induced *p53* mRNA destabilization (Fig. 4e; depletion of Gld4 reduced mostly cytoplasmic *p53* mRNA as examined by RT-PCR using exon-specific primers but had no effect on *p53* transcription as examined by RT-PCR using intron-specific primers).

The results presented here and in Katoh *et al.*¹⁶ suggest a model for homeostatic control of *p53* synthesis in human skin fibroblasts (Fig. 4f). Gld2 stabilizes miR-122 by catalysing the addition of a single adenylate residue to its 3' end¹⁶. miR-122 then base-pairs to two regions of the CPEB 3' UTR, causing instability and/or translational inhibition of the mRNA. CPEB, whose levels are modulated by these events, binds to the *p53* 3' UTR and recruits Gld4, which in turn

maintains *p53* mRNA polyadenylation and translation. We visualize this hierarchical regulation of *p53* to resemble a rheostat, where translation is turned up or down rather than a switch, where translation is turned on or off⁹, although *p53* mRNA translation can also be controlled by a switch mechanism in response to DNA damage^{20,21}. A 50% change in *p53* synthesis can toggle a cell between growth and senescence⁶, demonstrating that drastic biological consequences result from a relatively modest change in protein level.

Although ectopically expressed Gld2 immunoprecipitated from hepatocarcinoma cells adds a single adenosine to miR-122 *in vitro*¹⁶, Gld2 tethered to a small non-coding RNA by MS2 adds more than 300 adenylate residues in injected oocytes²², and about that same amount to mRNA when bound to CPEB¹⁰. How the enzyme can modulate its catalytic activity depending on the substrate is unknown, but we postulate that components of the RNA-induced silencing complex might be responsible. In addition to our demonstration that miR-122 is 3' mono-adenylated in skin fibroblasts, approximately 20% of all RNA deep sequencing reads from cloned neuroblastoma miRNAs contain a non-templated adenylate residue²³, suggesting that miR-122 may be one of several miRNAs that are mono-adenylated by Gld2.

In conclusion, our results demonstrate that Gld2 control of miR-122 stability in human skin fibroblasts tunes CPEB expression, which in turn regulates *p53* mRNA polyadenylation and translation by Gld4. The coordinated activities of these factors then gate entry into senescence. These studies also bring up two additional aspects of CPEB-related activities: how does Gld4, but not Gld2, associate with CPEB on *p53* mRNA, and what molecular machinery is responsible for miR-122 destruction upon Gld2 depletion? Deciphering the mechanisms involved would probably require analysis of the combinatorial associations of factors on different RNA substrates.

METHODS SUMMARY

Molecular biology and cell culture. Primary human foreskin fibroblasts obtained from the Cell Culture Core Facility of the Yale University Skin Disease Research Center were cultured as described²⁴ in Dulbecco's modified eagle's medium (DMEM) containing 10% fetal calf serum. Amphotropic retroviruses and shRNA-containing lentiviruses were produced by transient transfection of 293T cells with a transfer vector and amphotropic packaging plasmids encoding VSV-G and gag-pol using Lipofectamine 2000 (Invitrogen). Human cells at 50% confluency were infected for 8–12 h with viral supernatants containing 7 $\mu\text{g ml}^{-1}$ polybrene. Typically 70–90% infection efficiency was achieved as assessed by a green fluorescent protein (GFP)-encoding viral gene or by immunostaining with haemagglutinin (HA) antibody (Covance). After infection, fresh medium was added to the infected fibroblasts. Some cells were analysed by western blotting for *p53* (DO-1, Neomarkers) and β -actin (Abcam). Other cells were fixed with 0.2% glutaraldehyde and stained for β -galactosidase activity at acidic pH according to Dimri *et al.*²⁵.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 March 2010; accepted 4 February 2011.

Published online 10 April 2011.

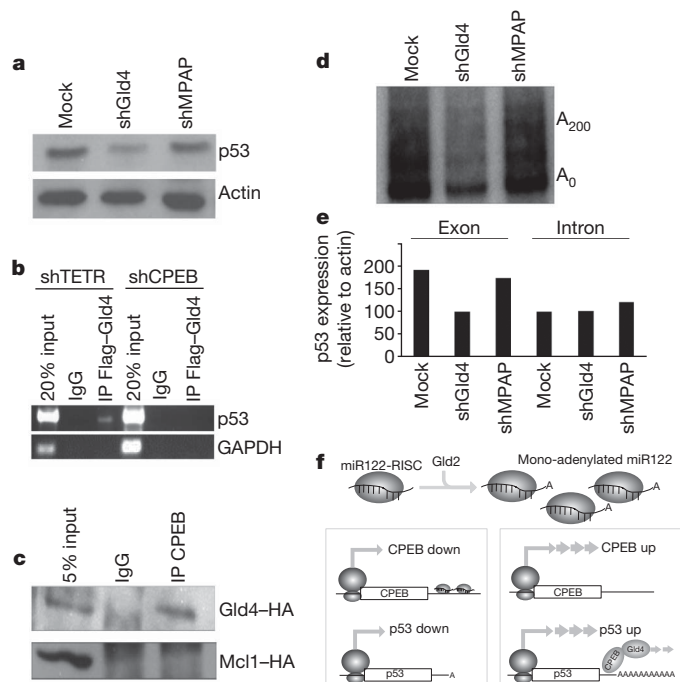


Figure 4 | Gld4 controls *p53* mRNA expression. **a**, *p53* and actin western blots from fibroblasts expressing GFP (Mock), shGld4 or shMitoPAP (shMitoPAP). **b**, Fibroblasts containing *shTETR* or *CPEB* were transfected with Gld4-Flag followed by Flag antibody or IgG immunoprecipitation of RNA complexes and RT-PCR for *p53* or GAPDH (control) RNAs. **c**, Protein from fibroblasts infected with Gld4-HA and CPEB-Flag was Flag or IgG immunoprecipitated and western blotted for HA. Other cells infected with CPEB-Flag and Mcl1-HA (a non-specific control) were processed similarly. **d**, Examination of *p53* poly(A) tail⁶ from skin fibroblasts expressing GFP or (Mock) Gld4 or mitoPAP shRNAs. **e**, RT-PCR analysis of predominantly cytoplasmic *p53* RNA (exon-specific primers), or nuclear *p53* pre-mRNA (intron-specific primers) in cells expressing GFP (Mock) or Gld4 or MitoPAP shRNAs. **f**, Model for regulation of *p53* translation; see text for explanation.

- Mendez, R. *et al.* Phosphorylation of CPE binding factor by Eg2 regulates translation of c-mos mRNA. *Nature* **404**, 302–307 (2000).
- Tay, J. & Richter, J. D. Germ cell differentiation and synaptonemal complex formation are disrupted in CPEB knockout mice. *Dev. Cell* **1**, 201–213 (2001).
- Alarcon, J. M. *et al.* Selective modulation of some forms of schaffer collateral-CA1 synaptic plasticity in mice with a disruption of the CPEB-1 gene. *Learn. Mem.* **11**, 318–327 (2004).
- Wu, L. *et al.* CPEB-mediated cytoplasmic polyadenylation and the regulation of experience-dependent translation of alpha-CaMKII mRNA at synapses. *Neuron* **21**, 1129–1139 (1998).
- Zearfoss, N. R., Alarcon, J. M., Trifilieff, P., Kandel, E. & Richter, J. D. A molecular circuit composed of CPEB-1 and c-Jun controls growth hormone-mediated synaptic plasticity in the mouse hippocampus. *J. Neurosci.* **28**, 8502–8509 (2008).
- Burns, D. M. & Richter, J. D. CPEB regulation of human cellular senescence, energy metabolism, and *p53* mRNA translation. *Genes Dev.* **22**, 3449–3460 (2008).
- Groisman, I. *et al.* Control of cellular senescence by CPEB. *Genes Dev.* **20**, 2701–2712 (2006).
- Campisi, J. & d'Adda di Fagagna, F. Cellular senescence: when bad things happen to good cells. *Nature Rev. Mol. Cell Biol.* **8**, 729–740 (2007).

9. Stewart, S. A. & Weinberg, R. A. Telomeres: cancer to human aging. *Annu. Rev. Cell Dev. Biol.* **22**, 531–557 (2006).
10. Barnard, D. C., Ryan, K., Manley, J. L. & Richter, J. D. Symplekin and xGLD-2 are required for CPEB-mediated cytoplasmic polyadenylation. *Cell* **119**, 641–651 (2004).
11. Kim, J. H. & Richter, J. D. Opposing polymerase-deadenylase activities regulate cytoplasmic polyadenylation. *Mol. Cell* **24**, 173–183 (2006).
12. Kim, J. H. & Richter, J. D. RINGO/cdk1 and CPEB mediate poly(A) tail stabilization and translational regulation by ePAB. *Genes Dev.* **21**, 2571–2579 (2007).
13. Lagos-Quintana, M. *et al.* Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**, 735–739 (2002).
14. Holst, L. M., Kaczowski, B. & Gniadecki, R. Reproducible pattern of microRNA in normal human skin. *Exp. Dermatol.* **19**, e201–e205 (2010).
15. Liao, J. Y. *et al.* Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3' trailers. *PLoS ONE* **5**, e10563 (2010).
16. Katoh, T. *et al.* Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. *Genes Dev.* **23**, 433–438 (2009).
17. Schmid, M., Kuchler, B. & Eckmann, C. R. Two conserved regulatory cytoplasmic poly(A) polymerases, GLD-4 and GLD-2, regulate meiotic progression in *C. elegans*. *Genes Dev.* **23**, 824–836 (2009).
18. Mullen, T. E. & Marzluff, W. F. Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. *Genes Dev.* **22**, 50–65 (2008).
19. Santos, S. D. & Ferrell, J. E. Systems biology: on the cell cycle and its switches. *Nature* **454**, 288–289 (2008).
20. Ofir-Rosenfeld, Y., Boggs, K., Michael, D., Kastan, M. B. & Oren, M. Mdm2 regulates p53 mRNA translation through inhibitory interactions with ribosomal protein L26. *Mol. Cell* **32**, 180–189 (2008).
21. Takagi, M., Absalon, M. J., McLure, K. G. & Kastan, M. B. Regulation of p53 translation and induction after DNA damage by ribosomal protein L26 and nucleolin. *Cell* **123**, 49–63 (2005).
22. Kwak, J. E., Wang, L., Ballantyne, S., Kimble, J. & Wickens, M. Mammalian GLD-2 homologs are poly(A) polymerases. *Proc. Natl Acad. Sci. USA* **101**, 4407–4412 (2004).
23. Schulte, J. H. *et al.* MicroRNAs in the pathogenesis of neuroblastoma. *Cancer Lett.* **274**, 10–15 (2009).
24. Rangarajan, A., Hong, S. J., Gifford, A. & Weinberg, R. A. Species- and cell type-specific requirements for cellular transformation. *Cancer Cell* **6**, 171–183 (2004).
25. Dimri, G. P. *et al.* A biomarker that identifies senescent human cells in culture and in aging skin *in vivo*. *Proc. Natl Acad. Sci. USA* **92**, 9363–9367 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank V. Ambros, J. Steitz, T. Kowalik and R. Davis for comments. S.N. was supported by a fellowship of the Max Planck Society and by European Molecular Biology Organization fellowship ALTF 995-2004. This work was supported by National Institutes of Health grant AG30323. Additional core support from the Diabetes Endocrinology Research Center (DK32520) is acknowledged.

Author Contributions Experiments were performed by D.B. and A.D. S.N. designed and constructed luciferase reporter constructs. D.B. and J.D.R. designed the experiments and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.D.R. (joel.richter@umassmed.edu).

METHODS

Analysis of p53, Gld2 and Gld4. Lentiviruses expressing shRNAs for Gld2, Gld4 and mitoPAP were generated as described⁶. shRNA against the TET repressor has also been described⁶. A retrovirus expressing Gld2-HA was generated as described⁷. Control and shCPEB infected fibroblasts were cultured in methionine and cysteine-free media (Invitrogen) for 45 min and then cultured in media containing 140 mCi [³⁵S]methionine and [³⁵S]cysteine (ProMix, Amersham) for 30 min. The cells were then washed and cultured in fresh DMEM supplemented with 2 mM each of methionine and cysteine for the times indicated. The cells were then frozen and stored until they were used for p53 immunoprecipitation and analysis by SDS-polyacrylamide gel electrophoresis and phosphorimaging. Cells were also cultured in methionine/cysteine-free media in the presence of MG132, a proteasome inhibitor, for 30 min followed by culture for 15 min in 100 μ Ci [³⁵S]methionine and cysteine; p53 was then immunoprecipitated and analysed as noted above.

Ligation-mediated polyadenylation test assays were used to detect the poly(A) tail of p53 mRNA in wild-type cells, shGld2 knockdown cells (two shRNAs targeting different regions of Gld2 were used), cells expressing ectopic CPEB and cells expressing ectopic that lacked a zinc finger and hence unable to bind RNA (CPEBAZF)⁶. Quantitative RT-PCR analyses were normalized against *actin* RNA.

Oxygen consumption and cellular senescence. To measure oxygen consumption, approximately 4×10^5 cells were washed and suspended in 200 ml Krebs-Ringer solution plus HEPES (125 mM NaCl, 1.4 mM KCl, 20 mM HEPES, pH 7.4, 5 mM NaHCO₃, 1.2 mM MgSO₄, 1.2 mM KH₂PO₄, 1 mM CaCl₂) containing 1% BSA. Cells from each condition were aliquoted into a BD Oxygen Biosensor System plate (BD Biosciences) in triplicate. Plates were assayed on a SAFIRE multimode microplate spectrophotometer (Tecan) at 1 min intervals for 60 min at an excitation wavelength of 485 nm and an emission wavelength of 630 nm.

Mock or shGld2-infected cells were stained for β -galactosidase at acidic pH, which denotes cellular senescence. Cell number was also determined with a haemocytometer; population doublings were plotted as growth curves of wild-type cells or cells infected with shGld2.

miR-122 cloning and sequencing. Small RNAs from human foreskin fibroblasts were extracted with mirVANA PARIS kit (Ambion). Those corresponding in length to 18–24 nucleotides were resolved by urea-polyacrylamide gel electrophoresis, extracted and ethanol precipitated. miRNA cloning linker-1 (IDT) was ligated to the 3' ends and used to prime a reverse transcription reaction with Superscript III

(Invitrogen) and the RT primer DP3 (5'-ATTGATGGTGCCTACAG-3'). cDNA was then PCR amplified with *miR-122* specific primer 5'-AGGGGCGCCTG GAGTGTGACAATG-3' and DP3. The PCR product was cloned into pGEM-T (Promega) and sequenced. The chromatogram shown is adapted from 4peaks (<http://mekentosj.com/science/4peaks/>).

Antagomir depletion of miR-122. LNA antagomir against *miR-122*, or a scrambled sequence LNA (Exiqon) were electroporated (Amaza, Lonza) at a final concentration of 4 nM into approximately 10^6 human foreskin fibroblast cells together with 0.8 μ g firefly (*pGL3*, Promega) and 1.0 μ g *Renilla* (*pRLTK*) luciferase-encoding plasmid, the latter harbouring the full-length CPEB 3' UTR or deletion mutations of the 3' UTR. Luciferase assays were performed 16–24 h after electroporation according to methods described elsewhere²⁶. The amount of *Renilla* luciferase activity derived from RNA containing the entire CPEB 3' UTR (full) was arbitrarily set at 100. When used, the CPEB 3' UTR deletions (Δ) were (in CPEB nucleotide number) as follows: Δ A, 420–755; Δ B, 480–755; Δ C, 530–755; Δ D, 565–755; Δ E, 640–755.

CPEB-Gld4 co-immunoprecipitation. Cells were transfected with plasmids encoding Flag-CPEB and HA-Gld4 using Effectene (Qiagen). The cells were then collected in PBS and lysed in lysis/wash buffer (50 mM Tris-HCl, pH 7.4, 100 mM NaCl, 1 mM MgCl₂, 0.1 mM CCl₂, 0.1% SDS and Complete Protease Inhibitor (Roche)). Extracts (0.5 mg protein) were incubated with M2-Flag antibody-(Sigma) coated Dynabeads (Invitrogen) for 2 h at 4 °C. The beads were then washed three times with lysis/wash buffer and the bound proteins eluted by boiling in SDS sample buffer. Co-immunoprecipitates were detected by western blotting with HA antibody (HA.11 16B12, Covance). Control immunoprecipitations were performed with generic mouse IgG-coated Dyanbeads.

Gld4-RNP co-immunoprecipitation. Mock or CPEB-depleted human foreskin fibroblasts⁶ were electroporated (Amaza, Lonza) with a plasmid encoding Flag-Gld4 according to the manufacturer's instructions. Immunoprecipitation with the Flag antibody followed the procedure of Peritz *et al.*²⁷ with the following modifications: (1) M2 anti-Flag (Sigma)-coated Dynabeads were used instead of agarose beads; and (2) washes with buffer containing 1 M urea were omitted. p53 RNA was detected by RT-PCR as described⁶.

26. Nottrott, S., Simard, M. J. & Richter, J. D. Human let-7a miRNA blocks protein production on actively translating polyribosomes. *Nature Struct. Mol. Biol.* **13**, 1108–1114 (2006).

27. Peritz, T. *et al.* Immunoprecipitation of mRNA-protein complexes. *Nature Protocols* **1**, 577–580 (2006).

Enzyme-catalysed [4+2] cycloaddition is a key step in the biosynthesis of spinosyn A

Hak Joong Kim¹, Mark W. Ruszczycky^{2*}, Sei-hyun Choi^{1*}, Yung-nan Liu² & Hung-wen Liu^{1,2}

The Diels–Alder reaction is a [4+2] cycloaddition reaction in which a cyclohexene ring is formed between a 1,3-diene and an electron-deficient alkene via a single pericyclic transition state¹. This reaction has been proposed as a key transformation in the biosynthesis of many cyclohexene-containing secondary metabolites^{2–5}. However, only four purified enzymes have thus far been implicated in biotransformations that are consistent with a Diels–Alder reaction, namely solanapyrone synthase⁶, LovB^{7,8}, macrophomate synthase^{9,10}, and riboflavin synthase^{11,12}. Although the stereochemical outcomes of these reactions indicate that the product formation could be enzyme-guided in each case, these enzymes typically demonstrate more than one catalytic activity, leaving their specific influence on the cycloaddition step uncertain. In our studies of the biosynthesis of spinosyn A, a tetracyclic polyketide-derived insecticide from *Saccharopolyspora spinosa*^{13,14}, we identified a cyclase, SpnF, that catalyses a transannular [4+2] cycloaddition to form the cyclohexene ring in spinosyn A. Kinetic analysis demonstrates that SpnF specifically accelerates the ring formation reaction with an estimated 500-fold rate enhancement. A second enzyme, SpnL, was also identified as responsible for the final cross-bridging step that completes the tetracyclic core of spinosyn A in a manner consistent with a Rauhut–Currier reaction¹⁵. This work is significant because SpnF represents the first example characterized *in vitro* of a stand-alone enzyme solely committed to the catalysis of a [4+2] cycloaddition reaction. In addition, the mode of formation of the complex perhydro-*as*-indacene moiety in spinosyn A is now fully established.

Spinosyn A (1), an active ingredient of several highly effective and environmentally benign commercial insecticides, has a complex aglycone structure comprising a perhydro-*as*-indacene moiety fused to a 12-membered macrolactone^{13,14}. How this tetracyclic ring system

is biosynthesized has been a subject of much speculation^{16–18}. Attention has largely focused on the construction of the cyclohexene ring due to the potential involvement of an enzyme that catalyses the [4+2] cycloaddition, which if concerted would represent a so-called ‘Diels–Alderase’. Four genes in the spinosyn A biosynthetic gene cluster of *S. spinosa*—*spnF*, *spnJ*, *spnL* and *spnM*—were proposed to convert product (2) of the polyketide synthase (PKS) to the tetracyclic aglycone (4) (see Fig. 1)¹⁶. The gene product of *spnJ*, which is a flavin-dependent dehydrogenase, was recently demonstrated to catalyse oxidation of the 15-OH of 2 to form the keto-intermediate 3 (ref. 19). However, the functions of the enzymes encoded by the remaining three genes, *spnF*, *spnL* and *spnM*, which show significant sequence similarity to lipases (SpnM) and *S*-adenosyl-L-methionine (SAM)-dependent methyltransferases (SpnF and SpnL)¹⁶, remain elusive.

To investigate the functions of SpnF, SpnL and SpnM, the corresponding genes were heterologously overexpressed in *Escherichia coli* BL21(DE3)* and their products purified as *N*-terminal His₆-tagged proteins (>95% purity). Each of the purified enzymes was tested for activity with 3. As shown in Fig. 2A, neither SpnF nor SpnL processes 3. In contrast, complete conversion of 3 to a new product was observed after a 2-h incubation with SpnM. NMR and mass spectrometry analysis of this new product (8) revealed a 15,6,5-tricyclic skeleton. Further investigation of the reaction time-course led to the discovery of a transient intermediate (Fig. 2D, orange peak), which was identified by spectral analysis as the monocyclic macrolactone 5.

These findings indicated that the SpnM-catalysed conversion of 3 to 8 might be a two-step process involving 1,4-dehydration of 3 followed by a transannular [4+2] cycloaddition between the $\Delta^{11,12}$ -alkene and the conjugated $\Delta^{4,5}$, $\Delta^{6,7}$ -diene of intermediate 5 to form the cyclohexene moiety in 8 (Fig. 4). To investigate whether SpnM is indeed a

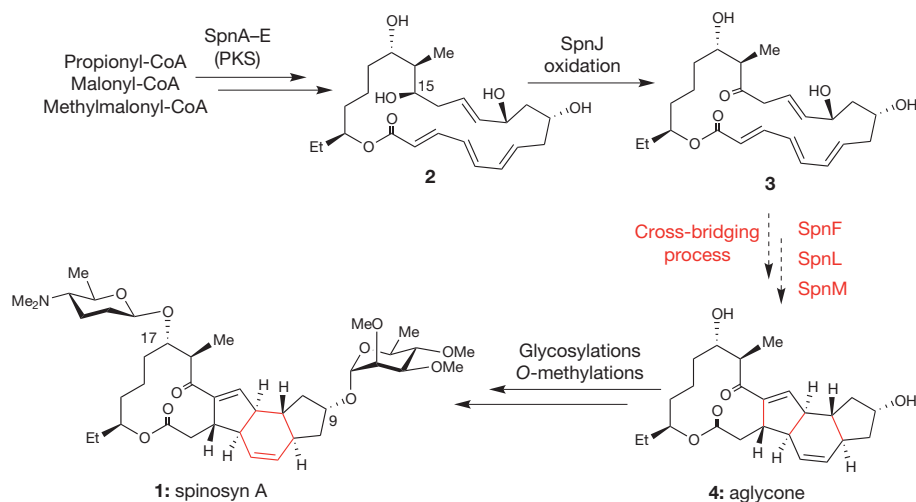


Figure 1 | Proposed spinosyn biosynthetic pathway.

¹Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, Texas 78712, USA. ²Division of Medicinal Chemistry, College of Pharmacy, University of Texas at Austin, Austin, Texas 78712, USA.

*These authors contributed equally to this work.

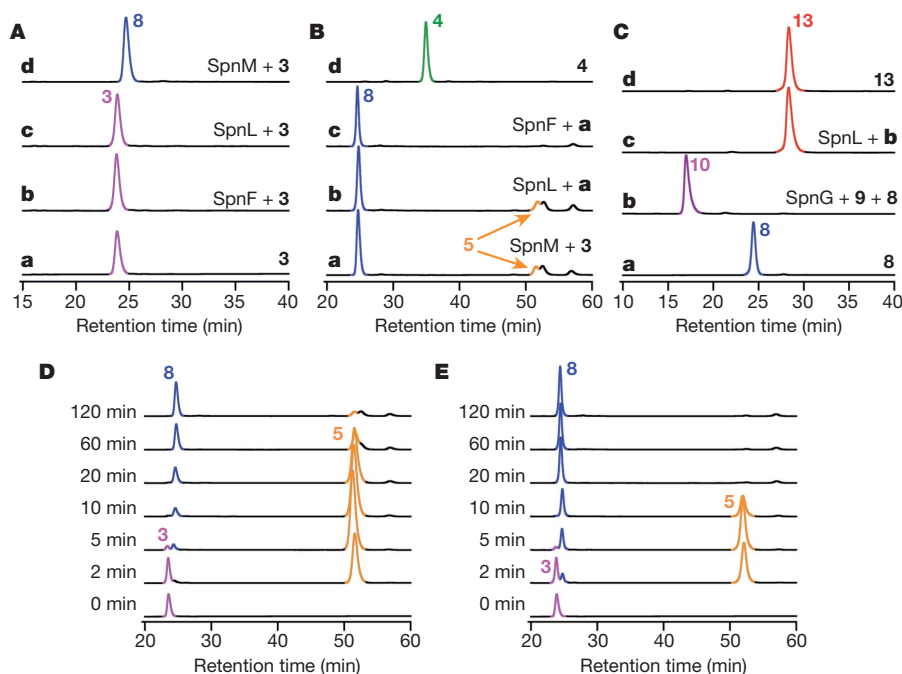


Figure 2 | HPLC analysis showing the reactions catalysed by SpnF, SpnL and SpnM. A, Incubation of 2 mM **3** for 2 h alone (a), and with either 1.25 μ M SpnF (b), 1.25 μ M SpnL (c), or 1.25 μ M SpnM (d). B, Incubation of 2 mM **8** (generated *in situ* from **3**) for 2 h either alone (a), with 15 μ M SpnL (b), or with 15 μ M SpnF (c). Trace (d) is the authentic standard of **4**. C, Incubation of

1 mM **8** for 2 h either alone (a); with both 20 μ M SpnG and 1.5 mM **9** (b); or with 15 μ M SpnL, 20 μ M SpnG and 1.5 mM **9** (c). Trace (d) is the authentic standard of **13**. D, Time-course of 2 mM **3** in the presence of 1.25 μ M SpnM. E, Time course of 2 mM **3** in the presence of both 1.25 μ M SpnM and 10 μ M SpnF.

cyclase of dual function catalysing both the dehydration and cyclization steps, the dependence of the rate of each step on the concentration of SpnM was determined. In this experiment the formation and consumption of **5**, which respectively reflect the dehydration and cyclization steps, was monitored by high-performance liquid chromatography (HPLC) as a function of time.

As shown in Fig. 3a, rate enhancement of the dehydration step was observed with the increase in SpnM concentration, whereas the rate of cyclization was unaffected. The full time courses of the formation and decay of **5** were analysed using numerically integrated coupled rate equations based on different kinetic models (see Supplementary Information Section 5)²⁰. The observed data was fit best to the integrated rate equations (1a) and (1b), which model Michaelis–Menten kinetics for the dehydration step (V_{SpnM} and K_{SpnM}) and first order

kinetics (k_{non}) for the cyclization step. Correlation of the fitted parameters versus the concentration of SpnM (Fig. 3b) reveals a significant dependence of V_{SpnM} , but not K_{SpnM} or k_{non} , on SpnM concentration. This analysis indicates that SpnM functions only as a dehydratase, and its product (**5**) can cyclize nonenzymatically to **8** with a first order rate constant of approximately 0.03 min^{-1} .

Having established that SpnM functions as a dehydratase, whereas the [4+2] cycloaddition can proceed nonenzymatically, we next considered the possible involvement of the two remaining gene products,

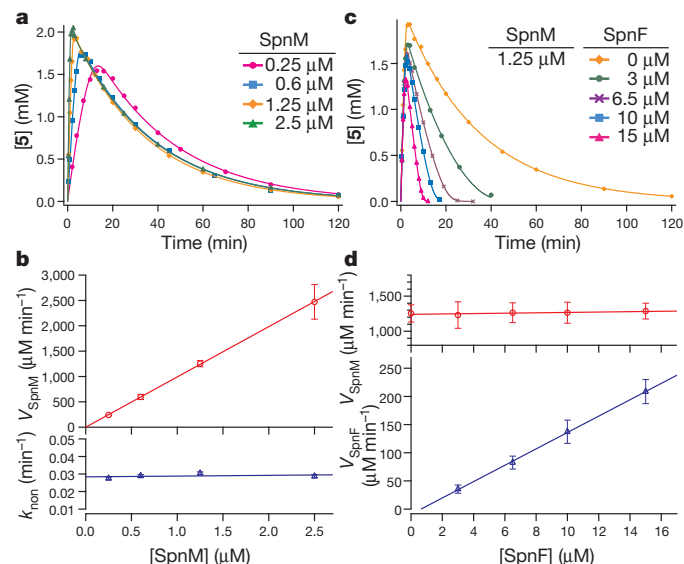


Figure 3 | Kinetic analysis demonstrating that SpnM and SpnF, respectively, catalyse the dehydration and cyclization steps of macrolactone 3. a, c, Formation and consumption of **5** was monitored by HPLC and the concentration (**5**) plotted versus time. Each reaction mixture initially contained 2 mM **3**, and the indicated amounts of SpnM and SpnF. a, Variable SpnM with no SpnF and fits based on the rate equations (1).

$$\frac{d[3]}{dt} = -\frac{V_{\text{SpnM}}[3]}{K_{\text{SpnM}} + [3]} \quad (1a)$$

$$\frac{d[5]}{dt} = \frac{V_{\text{SpnM}}[3]}{K_{\text{SpnM}} + [3]} - k_{\text{non}}[5] \quad (1b)$$

b, Fitted parameters V_{SpnM} and k_{non} versus $[SpnM]$. The measured turnover number for SpnM, $k_{\text{cat,SpnM}}$, is $1,020 \pm 57 \text{ min}^{-1}$ whereas the first order rate constant for nonenzymatic cyclization of **5**, k_{non} , is $0.0288 \pm 0.00041 \text{ min}^{-1}$. The apparent Michaelis constant for SpnM, K_{SpnM} , is $380 \pm 51 \mu\text{M}$. c, Variable SpnF at a fixed concentration of SpnM and fits based on rate equations (2).

$$\frac{d[3]}{dt} = -\frac{V_{\text{SpnM}}[3]}{K_{\text{SpnM}} + [3]} \quad (2a)$$

$$\frac{d[5]}{dt} = \frac{V_{\text{SpnM}}[3]}{K_{\text{SpnM}} + [3]} - k_{\text{non}}[5] - \frac{V_{\text{SpnF}}[5]}{K_{\text{SpnF}} + [5]} \quad (2b)$$

d, Correlation of V_{SpnM} and V_{SpnF} versus $[SpnF]$. The turnover number for SpnF, $k_{\text{cat,SpnF}}$, is $14 \pm 1.6 \text{ min}^{-1}$. The apparent Michaelis constant for SpnF, K_{SpnF} , which is independent of SpnF concentration, is $120 \pm 46 \mu\text{M}$. All values and error bars are \pm s.e.

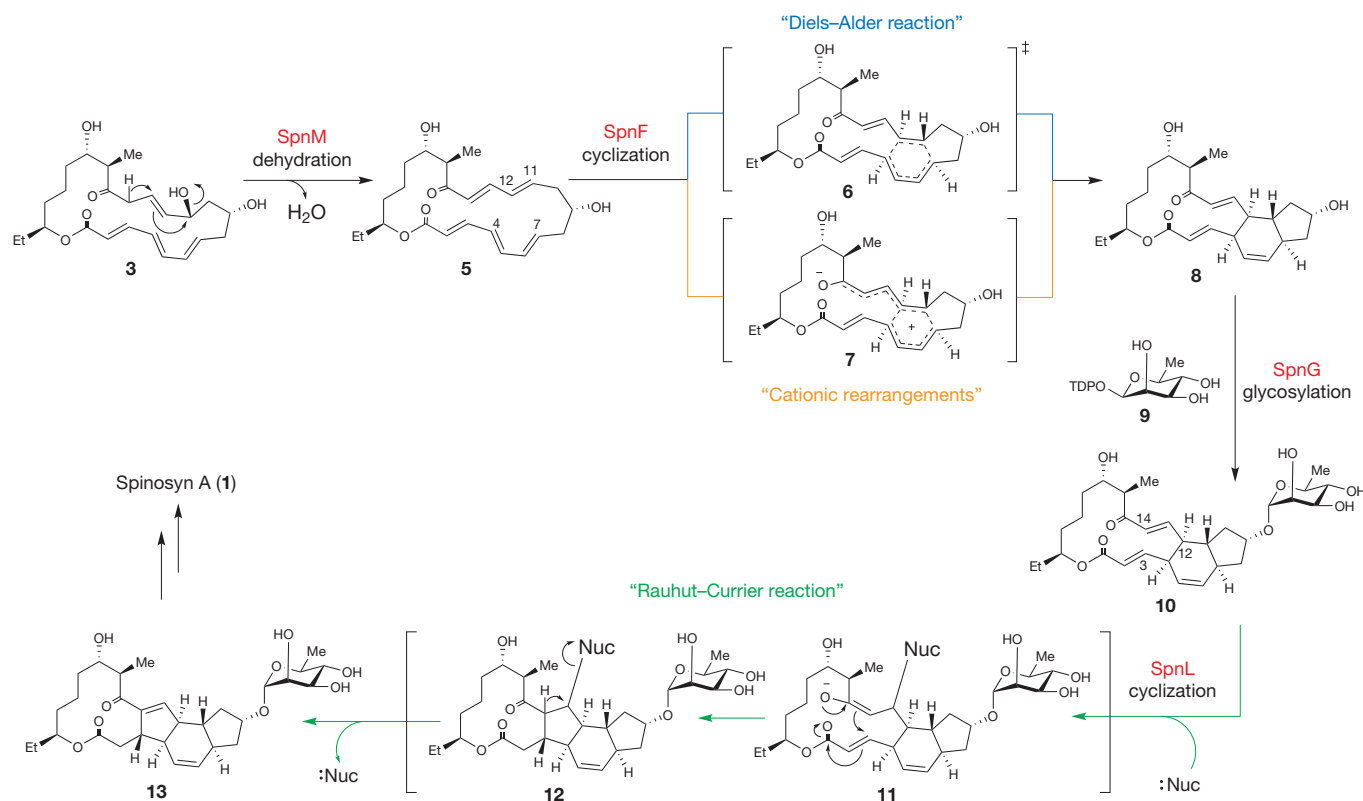


Figure 4 | The spinosyn aglycone biosynthetic pathway. SpnM catalyses a dehydration reaction to convert **3** to **5**, and SpnF subsequently catalyses cyclization of **5** to afford **8**. The resulting tricyclic macrolactone **8** is then modified with a rhamnose moiety at the C-9 hydroxyl group by SpnG rather

than being converted directly to the aglycone **4** (see Fig. 1) as previously thought. SpnL completes the cross-bridging process by interlinking the C-3 and C-14 carbon centres of **10** to produce the tetracyclic nucleus (**13**) of spinosyn A.

SpnF and SpnL, in the final C-C bond formation between C-3 and C-14. Surprisingly, incubation of SpnF or SpnL with **8**, which was generated *in situ* from **3** in the presence of SpnM, did not produce the aglycone **4** (Fig. 2B). However, when SpnF was added to the assay mixture, a change in the product distribution was noted including rapid disappearance of **5** (Fig. 2B, trace c). As shown in Fig. 2E, the consumption of **5** with concomitant formation of a product having a retention time consistent with **8** was completed in 20 min in the presence of 10 μM SpnF, instead of requiring more than 2 h in its absence (Fig. 2D). The structure of this product was established to be **8** by NMR analysis. Additional controls, including examining the effects of denatured SpnF as well as mutated SpnF on the rate of cyclization, were also performed. In all cases only native (His₆-tagged) SpnF led to appreciably increased rates of cyclization (see Supplementary Information Section 4.4). These results clearly indicate that SpnF is responsible for the observed rate enhancement of the cyclization of **5** to **8**.

To quantify the effect of SpnF on the rate of the cyclization step, the production and consumption of **5** was again monitored by HPLC, this time as a function of SpnF concentration while keeping that of SpnM fixed. As shown in Fig. 3c, the rate of conversion of **5** to **8** is clearly enhanced as the concentration of SpnF increases. The data was fit best using the coupled rate equations (2a) and (2b), where the cyclization event is modelled as the sum of a first order and a Michaelis–Menten process. In these fits all parameters including the initial concentration of **3** were allowed to float except for k_{non} , which was fixed at the fitted value obtained in the absence of SpnF. Correlation of the fitted parameters with SpnF concentration indicates a significant dependence of V_{SpnF} on SpnF concentration (Fig. 3d), which is not true for the V_{SpnM} parameter. These results firmly establish that SpnF is a cyclase catalysing the conversion of **5** to **8** with an apparent k_{cat} of $14 \pm 1.6 \text{ min}^{-1}$ (\pm s.e.) for an estimated rate enhancement ($k_{\text{cat,SpnF}}$ versus k_{non}) of approximately 500-fold.

This result left *spnL* as the only gene without an assigned function, and it was proposed that SpnL has a role in the transannular cyclization reaction between C-3 and C-14. However, the observed inability of SpnL to convert **8** to the anticipated product (**4**) (Fig. 2B, trace b) prompted us to reconsider the sequence of events of the proposed biosynthetic pathway in Fig. 1. Previously, glycosylations had been thought to occur on the tetracyclic aglycone (**4**), because transfer of the rhamnose moiety from thymidine diphosphate- β -L-rhamnose (TDP- β -L-rhamnose, **9**) to 9-OH of **4** by the rhamnosyltransferase, SpnG, had been demonstrated^{21,22}. However, this observation did not necessarily rule out an alternative pathway in which rhamnosylation of **8** precedes C-3/C-14 cross-bridging, because glycosyltransferases involved in the biosynthesis of secondary metabolites are known to be promiscuous with regard to their substrate specificity^{23,24}.

To test this possibility, the ability of SpnG to accept **8** as a substrate was investigated. As shown in Fig. 2C (trace b), incubation of **8** and **9** with SpnG resulted in the disappearance of **8** with the concomitant appearance of a new peak (retention time of 17 min). This new product was identified as **10** by NMR and mass spectrometry analysis (Fig. 4). Upon addition of SpnL to this reaction mixture, the peak at 17 min disappeared and a new peak appeared at 27 min (Fig. 2C, trace c). Both transformations were highly efficient. Identification of the new product as **13** was confirmed by high-resolution mass spectrometry analysis and HPLC co-elution with an authentic sample of **13**. These results strongly suggest that **10** rather than **8** is the biological substrate for SpnL, which catalyses the final cyclization step to generate the perhydro-*as*-indacene core.

The mechanisms by which SpnF and SpnL catalyse their respective cyclization reactions are a point of interest. The SpnF-catalysed *endo*-mode *syn*-addition of an alkenyl to a dienyl functionality seems consistent with a Diels–Alder reaction; however, confirmation of this hypothesis will require demonstrating that the reaction progresses

through a single pericyclic transition state such as **6**. Therefore, a stepwise [4+2] cycloaddition mechanism, for example, one involving a dipolar intermediate such as **7**, cannot at present be ruled out. In contrast, the C-C bond formation catalysed by SpnL may involve a Rauhut–Currier mechanism¹⁵ consistent with the observation that **10** is susceptible towards nucleophilic addition by a thiol (see Supplementary Information Section 3.10) forming a covalent adduct that may be structurally analogous to **11** or **12** (see Fig. 4), although the specific site of attack remains unknown. Whereas these mechanistic proposals are at present speculative, it is worth noting that Roush and co-workers were able to accomplish their non-enzymatic total synthesis of spinosyn A by exploiting both the transannular Diels–Alder and Rauhut–Currier reactions in an analogous fashion²⁵. This precedent in chemical synthesis certainly substantiates the feasibility of the mechanisms proposed for the SpnF- and SpnL-catalysed reactions.

In summary, the biosynthetic pathway for spinosyn A is now fully established (Fig. 4), with the specific functions of SpnM as a dehydratase and SpnF as well as SpnL as the two cyclases in the cross-bridging steps biochemically determined. SpnF represents the first enzyme for which specific acceleration of a [4+2] cycloaddition reaction has been experimentally verified as its only known function. It stands in contrast to macrophomate synthase, for which evidence has been provided suggesting a tandem Michael–aldol reaction mechanism^{26,27}, as well as the multifunctional solanapyrone synthase, LovB and riboflavin synthase, which participate in hydroxyl oxidation^{5,6}, polyketide synthesis^{7,8}, and hydride transfer¹², respectively, in addition to the [4+2] cycloaddition reactions, the concertedness of which have yet to be verified. For this reason, the SpnF reaction provides a unique system for detailed mechanistic investigation of enzyme-catalysed [4+2] cycloadditions and the existence of a bona fide Diels–Alderase.

METHODS SUMMARY

All proteins used in this work were overexpressed in *E. coli* BL21(DE3)* (Invitrogen) and purified by Ni-NTA (Qiagen) affinity chromatography. Specifically, SpnF was co-overexpressed with the chaperone protein pair, GroEL/ES, to improve its solubility. Because overexpression of the protein encoded by the originally assigned *spnM* gene¹⁶ failed to afford an active soluble protein, the gene sequence was re-examined and revised to include 204 additional base pairs (Supplementary Fig. 2). Overexpression of the revised *spnM* gene produced an active enzyme with significantly improved protein yield. All enzyme reaction products (**5**, **8**, **10** and **13**) were extracted with ethyl acetate or chloroform and purified using silica gel column chromatography or HPLC. Their structures were characterized by 1D- and 2D-NMR spectroscopy and/or high-resolution mass spectrometry. In particular, the stereochemistry of **8** was assigned based on its ¹H-¹H nuclear Overhauser enhancement spectroscopy (NOESY) spectrum. All substrate specificity and time-course assays were run in 50 mM Tris-HCl buffer (pH 8) at 30 °C. Reaction aliquots were quenched with an excess volume of ethanol after a given incubation time and centrifuged to remove protein. The supernatant was then analysed by reverse phase HPLC with detection by ultraviolet absorbance at 254 nm (Fig. 2) or 280 nm (Fig. 3). Time course assays also included *p*-methoxyacetophenone as an internal standard to normalize the peak areas corresponding to **5**. Numerical integration of equations (1) and (2) used the fourth order Runge–Kutta algorithm²⁸ following non-dimensionalization of substrate concentration. The resulting simulated progress curves were fit using steepest descent²⁹ directly to full time-courses of normalized substrate concentration obtained via the HPLC discontinuous assay to provide the kinetic parameters and a concentration normalization factor. Further detail regarding experimental procedures as well as data fitting and analysis is described in the Supplementary Information.

Received 29 March 2010; accepted 3 March 2011.

- Huisgen, R. Cycloadditions — definition, classification, and characterization. *Angew. Chem. Int. Edn Engl.* **7**, 321–328 (1968).
- Stocking, E. M. & Williams, R. M. Chemistry and biology of biosynthetic Diels–Alder reactions. *Angew. Chem. Int. Ed.* **42**, 3078–3115 (2003).
- Oikawa, H. Involvement of the Diels–Alderase in the biosynthesis of natural products. *Bull. Chem. Soc. Jpn.* **78**, 537–554 (2005).
- Kelly, W. L. Intramolecular cyclizations of polyketide biosynthesis: mining for a “Diels–Alderase?”. *Org. Biomol. Chem.* **6**, 4483–4493 (2008).

- Oikawa, H. in *Comprehensive Natural Products II, Chemistry and Biology* (eds Mander, L. & Liu, H.-w.) Vol. 8 277–314 (Elsevier, 2010).
- Oikawa, H., Katayama, K., Suzuki, Y. & Ichihara, A. Enzymatic activity catalyzing exo-selective Diels–Alder reaction in solanapyrone biosynthesis. *J. Chem. Soc. Chem. Comm.* 1321–1322 (1995).
- Auclair, K. et al. Lovastatin nonaketide synthase catalyzes an intramolecular Diels–Alder reaction of a substrate analogue. *J. Am. Chem. Soc.* **122**, 11519–11520 (2000).
- Ma, S. M. et al. Complete reconstitution of a highly reducing iterative polyketide synthase. *Science* **326**, 589–592 (2009).
- Watanabe, K., Mie, T., Ichihara, A., Oikawa, H. & Honma, M. Detailed reaction mechanism of macrophomate synthase. *J. Biol. Chem.* **275**, 38393–38401 (2000).
- Ose, T. et al. Insight into a natural Diels–Alder reaction from the structure of macrophomate synthase. *Nature* **422**, 185–189 (2003).
- Eberhardt, S. et al. Domain structure of riboflavin synthase. *Eur. J. Biochem.* **268**, 4315–4323 (2001).
- Kim, R.-R. et al. Mechanistic insights on riboflavin synthase inspired by selective binding of the 6,7-dimethyl-8-ribityllumazine exomethylene anion. *J. Am. Chem. Soc.* **132**, 2983–2990 (2010).
- Kirst, H. A. et al. A83543A-D, Unique fermentation-derived tetracyclic macrolides. *Tetrahedr. Lett.* **32**, 4839–4842 (1991).
- Kirst, H. A. The spinosyn family of insecticides: realizing the potential of natural products research. *J. Antibiot.* **63**, 101–111 (2010).
- Aroyan, C. E., Dermenci, A. & Miller, S. J. The Rauhut–Currier reaction: a history and its synthetic application. *Tetrahedron* **65**, 4069–4084 (2009).
- Waldron, C. et al. Cloning and analysis of the spinosad biosynthetic gene cluster of *Saccharopolyspora spinosa*. *Chem. Biol.* **8**, 487–499 (2001).
- Martin, C. J. et al. Heterologous expression in *Saccharopolyspora erythraea* of a pentaketide synthase derived from the spinosyn polyketide synthase. *Org. Biomol. Chem.* **1**, 4144–4147 (2003).
- Oikawa, H. Biosynthesis of structurally unique fungal metabolite GKK1032A₂: indication of novel carbocyclic formation mechanism in polyketide biosynthesis. *J. Org. Chem.* **68**, 3552–3557 (2003).
- Kim, H. J., Pongdee, R., Wu, Q., Hong, L. & Liu, H.-w. The biosynthesis of spinosyn in *Saccharopolyspora spinosa*: synthesis of the cross-bridging precursor and identification of the function of SpnJ. *J. Am. Chem. Soc.* **129**, 14582–14584 (2007).
- Frieden, C. Analysis of kinetic data: practical applications of computer simulation and fitting programs. *Methods Enzymol.* **240**, 311–322 (1994).
- Chen, Y., Lin, Y., Tsai, K. & Chiu, H. Functional characterization and substrate specificity of spinosyn rhamnosyltransferase by *in vitro* reconstitution of spinosyn biosynthetic enzymes. *J. Biol. Chem.* **284**, 7352–7363 (2009).
- Kim, H. J. et al. Biosynthesis of spinosyn in *Saccharopolyspora spinosa*: synthesis of permethylated rhamnose and characterization of the functions of SpnH, SpnI, and SpnK. *J. Am. Chem. Soc.* **132**, 2901–2903 (2010).
- Thibodeaux, C. J., Melançon, C. E. & Liu, H.-w. Unusual sugar biosynthesis and natural product glycodiversification. *Nature* **446**, 1008–1016 (2007).
- Thibodeaux, C. J., Melançon, C. E. & Liu, H.-w. Natural-product sugar biosynthesis and enzymatic glycodiversification. *Angew. Chem. Int. Ed.* **47**, 9814–9859 (2008).
- Mergott, D. J., Frank, S. A. & Roush, W. R. Total synthesis of (–)-spinosyn A. *Proc. Natl Acad. Sci. USA* **101**, 11955–11959 (2004).
- Guimarães, C. R. W., Udier-Blagović, M. & Jørgensen, W. L. Macrophomate synthase: QM/MM simulations address the Diels–Alder versus Michael–aldol reaction mechanism. *J. Am. Chem. Soc.* **127**, 3577–3588 (2005).
- Serafimov, J. M., Gillingham, D., Kuster, S. & Hilvert, D. The putative Diels–Alderase macrophomate synthase is an efficient aldolase. *J. Am. Chem. Soc.* **130**, 7798–7799 (2008).
- Tenenbaum, M. & Pollard, H. *Ordinary Differential Equations* (Dover, 1985).
- Draper, N. R. & Smith, H. *Applied Regression Analysis* 3rd edn (Wiley-Interscience, 1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Whitman for review of the manuscript, L. Hong for carrying out the early cloning work, B. Shoulders and S. Sorey for assistance with the interpretation of the NMR spectra, and S. Mansoorabadi and E. Isiorho for discussions on the reaction mechanisms and structural modelling of SpnF. This work is supported in part by grants from the National Institutes of Health (GM035906, F32AI082906), the Texas Higher Education Coordination Board (ARP-003658-0093-2007), and the Welch Foundation (F-1511).

Author Contributions H.-w.L. provided the scientific direction and the overall experimental design for the studies. H.J.K. designed and performed most of the experiments. S.-h.C. participated in the chemical synthesis of the substrate for SpnJ (**2**) and the characterization of the structures of the enzyme reaction products. M.W.R. analysed the kinetic experiment data. Y.-n.L. carried out the mutation studies of SpnF. H.J.K., M.W.R. and H.-w.L. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.-w.L. (h.w.liu@mail.utexas.edu).

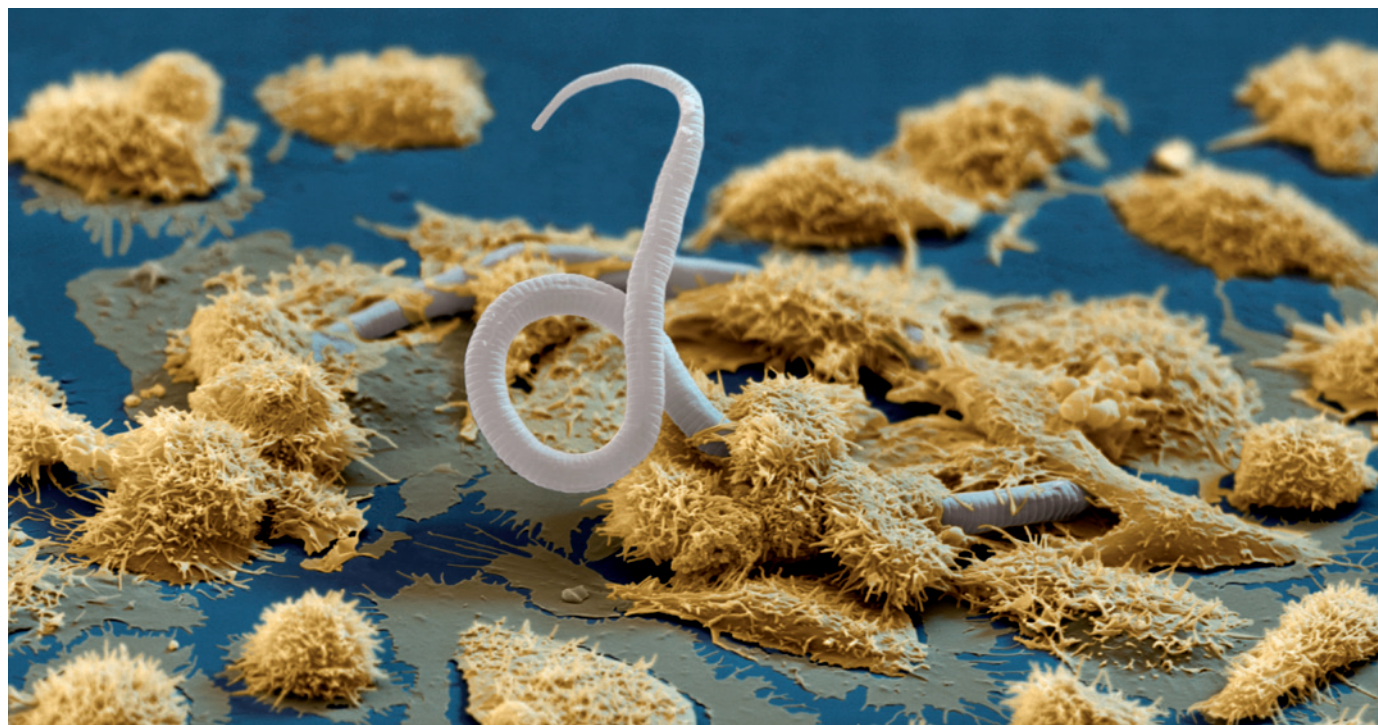
CAREERS

TURNING POINT Biologist aims to advance understanding of photosynthesis **p.115**

AGRICULTURE Increased budgets create research jobs in sub-Saharan Africa **p.115**

NATUREJOBS For the latest career listings and advice www.naturejobs.com

EYE OF SCIENCE/SPL



Systems immunologists track the effects of drugs and infectious agents, such as parasitic worms, on the body's defences.

SYSTEMS IMMUNOLOGY

Complexity captured

Researchers who can grasp the intricacies of the immune system and enjoy distilling meaning from large data sets are in demand for a growing subfield of systems biology.

BY CHARLOTTE SCHUBERT

Lisa Israelsson grips her head in her hands, pink hair peeking through her fingers. She has hit the wrong key on her computer, and now she can't find the 1,511,104 data points that she had been working on. But that's the least of her worries — once she has retrieved the information, she will have to work out what it all means.

The numbers are ranged on a spreadsheet with 47,222 rows and 32 columns. Each column corresponds to a blood sample taken from a person with an immune-system defect. Every sample has been processed and placed onto a microarray chip dotted with 47,222 unique DNA molecules to measure which genes in the sample are active, which

are inactive and which are expressed at unusual levels. It is going to take Israelsson, an immunology postdoc with little experience in computational biology, about a week to crunch the information and turn it into tidy charts that will provide quick snapshots of the state of each person's immune system.

Fortunately, Israelsson can turn for help to any of the researchers in her lab, who include molecular biologists, biostatisticians, bioinformaticians and a software engineer who once designed an electronic billing system for the government of Bhutan. The lab, at the Benaroya Research Institute in Seattle, Washington, is run by Damien Chaussabel, a systems immunologist. Using readouts of the immune system such as the one that Israelsson is preparing, the lab aims to develop molecular signatures

for immune-related conditions from sepsis to lupus, to better diagnose patients and predict how they will respond to drugs and vaccines.

Chaussabel's cooperative, interdisciplinary team is typical in the burgeoning field of systems immunology, a loosely defined subspecialty of systems biology. As the name implies, systems biology measures how molecular components such as genes, proteins and cells interact within a system — from a biological tissue to an entire organism — and uses computational mathematical methods to describe the system and predict how it will behave if its components are perturbed.

Until a few years ago, systems biology was dominated by the study of yeast cells and tumours, which have a manageable number of components. The immune system, with its ►

► dozens of different cell types and hundreds of intersecting molecular pathways and signals, has proved more difficult to model. But now, researchers are taking up the challenge.

A systems approach can help immunologists to predict, for example, how the immune system will respond to a particular vaccine or drug — will it react in a way that will ease disease? Or will a drug that is under development cause undesirable side effects?

Encouraged by early successes of systems-biology approaches in cancer research — including the commercialization of diagnostic gene chips — and by technological advances in the large-scale analyses of molecules, systems immunology is in a growth stage: funding is expanding, creating job opportunities for researchers at all stages of their careers. Although much of the demand in academia is coming from the United States, there are also opportunities in Europe and Asia. And pharmaceutical companies worldwide are hiring systems immunologists at all levels to help pick out potential drug targets on the basis of models of immune-system function, or to monitor the immune systems of subjects in clinical trials.

GROWING SUPPORT

In Seattle, not far from where Israelsson is working on her data set, Alan Aderem's 45-member systems-immunology lab is in the middle of moving from the Institute of Systems Biology, which Aderem co-founded, to the Seattle Biomedical Research Institute (SBRI), which he will now head. The move shows how interest in the field of systems immunology is growing. It was financed by a US\$7-million grant from the Bill & Melinda Gates Foundation in Seattle, which, among other work, supports research into infectious diseases that affect the developing world. Aderem's lab will bring a systems approach to the SBRI's current focus on infectious diseases, including HIV, malaria, tuberculosis and leishmaniasis.

Funding agencies in the United States are helping to propel the field's expansion. Chaussabel's research is funded largely by a five-year, \$100-million US National Institutes of Health (NIH) grant to the Baylor Institute for Immunology Research in Dallas, Texas — where Chaussabel was based until last year and still holds a part-time appointment — and six other centres nationwide that are part of the NIH's Human Immunology Project Consortium. The consortium aims to generate molecular signatures of the human immune system and its response to vaccination and infection.

The NIH also spends millions of dollars on systems immunology outside the consortium. For example, the largest formal systems-biology programme on the NIH campus is in systems immunology, says Ron Germain, an immunologist at the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland, who launched the scheme in 2008. His programme had a budget of \$4.1 million

last year, is hiring a fifth principal investigator, and will employ dozens of postdocs, technicians and other staff as it develops.

But Chaussabel cautions that as with any new discipline, systems immunology will continue to grow only if it proves successful; for example, if researchers can come up with improved approaches to HIV-vaccine development or new diagnostic tests for a variety of diseases. "You need to see some deliverables in the short- to mid-term before the field gets very large," he says. However, Chaussabel, Aderem and others in the field are optimistic about prospects for systems immunologists in academia and industry. "In an environment where all jobs are really tight," says Aderem, "systems biologists, the good ones, have their pick to a large extent."

GLOBAL PROSPECTS

Major funding agencies outside the United States have not embraced systems immunology with as much enthusiasm as the NIH. But projects are starting to emerge worldwide.

Last year, for instance, John Connolly of the Singapore Immunology Network started a systems-biology lab that focused on the

human immune system. His group has big ambitions: he aims to collaborate with drug companies and researchers throughout Asia to monitor the immune responses of subjects in clinical trials and identify molecular signatures that predict outcomes or side effects.

The European Union (EU) has not

funded any efforts comparable to the Human Immunology Project Consortium, says Stefan Kaufmann, director of immunology at the Max Planck Institute for Infection Biology in Berlin. But EU researchers can obtain grants for smaller systems-immunology projects. In Kaufmann's lab, for example, a study of tuberculosis is funded in part by SystemeTb, a four-year, €10.5-million (US\$15.2-million) project launched last year and financed by the EU's research-funding framework, which brings together systems biologists and tuberculosis experts at 13 institutions across Europe.

Academia is not the only area in which opportunities are growing. The pharmaceutical industry worldwide is hiring systems immunologists, from newly minted PhDs in biostatistics, bioinformatics, pharmacodynamics and other computational fields to immunologists and molecular biologists who have spent years as postdocs or academic faculty members.

The increased demand in recent years has coincided with the overall contraction of

pharmaceutical research budgets, so it is difficult to ascertain whether there is an overall uptick in industry job openings, warns Debraj GuhaThakurta, who heads the systems-immunology group at Dendreon, a biotechnology company based in Seattle.

The greatest demand for systems immunologists in industry is at large pharmaceutical companies that are developing vaccines or other agents that affect the immune system, such as monoclonal antibodies, says Aderem. Those companies include Novartis, based in Basel, Switzerland, and Sanofi-aventis, headquartered in Paris.

WHAT IT TAKES TO GET THE JOB

Those interested in a career in systems immunology should get experience in both systems biology and immunology, says Aderem. He advises immunology graduate students to spend some time in systems-biology labs, and quantitative experts, such as biostatisticians and bioinformaticians, to learn some immunology. GuhaThakurta, who has a PhD in biophysics, boned up on immunology through reading and coursework for six months before applying to Dendreon. Institutions such as the University of California, Irvine, are also starting to offer graduate programmes that combine specific fields of biology with training in systems biology.

In comparison with other fields, systems-biology labs hire a lot of staff scientists, who have greater job security and higher salaries than postdocs. Aderem's lab, for instance, employs 13 research scientists, a set-up that aims to foster long-term collaboration.

Applicants will find that the abilities to work in a team and think outside their areas of training are important in this multidisciplinary area of research. Such qualities are particularly valued in industry, where systems immunologists interact on a daily basis with chemists, pharmacologists and clinical-trial designers. Industry also seems to place a premium on scientists with experience in translational research, says GuhaThakurta.

Many biologists are intimidated by the computational side of science, and computer scientists can be perplexed by biological questions, but to be successful in this field "both groups need to be prepared to get out of their comfort zones," says Aderem.

It is just that adventurous quality that led Israelsson to Chaussabel's lab and allowed her to tackle systems immunology. Israelsson and a colleague, a computational biologist, are now coaxing her data out of hiding. When they have finished, Israelsson will learn how to use software that tells her which genes in the people in her study are going rogue. Her work, says Israelsson, is never boring. "Every day," she says, "there is something new." ■

Charlotte Schubert is a freelance journalist in Seattle, Washington.



"Every day there is something new."

Lisa Israelsson

C. SCHUBERT

TURNING POINT

Martin Jonikas

Martin Jonikas, a plant biologist at the Carnegie Institution for Science in Stanford, California, won one of four grants for research to increase the efficiency of photosynthesis, awarded jointly on 28 March by the US National Science Foundation (NSF) and the UK Biotechnology and Biological Science Research Council (BBSRC).

How did you become interested in biology?

During my undergraduate degree in aerospace engineering at the Massachusetts Institute of Technology in Cambridge, I took a required course in molecular biology. Biological machines can make complex proteins that humans can't, and I thought that biology was going to become a major frontier for engineering. I wanted to be part of it.

Did you go straight into plant biology?

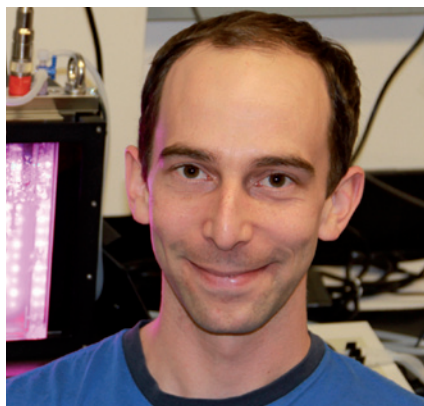
No, I did a PhD at the University of California, San Francisco, on basic molecular biology. We used genome-wide screening to identify a new pathway required for protein folding in yeast. While planning research proposals to apply for positions as a fellow, I realized that no one was applying high-throughput genetic tools to photosynthesis, one of my areas of interest, and that I could fill that niche. The time spent refining proposals helped me to secure a position as a 'staff associate' at the Carnegie Institution for Science.

How is that different from a normal postdoc?

I have a five-year non-tenure-track position. I'll be working to characterize the genes at work in photosynthesis and to make it more efficient. One of the benefits is that I can run a lab and assemble a team, so we can work on problems in depth and do exciting research.

How did you get your grant?

It was an unusual process. Last September, the NSF and the BBSRC assembled about 30 researchers to brainstorm on how to improve photosynthesis. I thought that this was a wonderful opportunity, and a grant would just be the cherry on top. One of the constraints on photosynthesis is that the primary enzyme that converts carbon dioxide into sugars, Rubisco, works best under higher carbon dioxide levels than exist in the atmosphere today. But some plants can concentrate carbon dioxide around Rubisco. Three colleagues and I suggested characterizing the components of this concentration mechanism, and trying to put them into some crop plants. We're not claiming that we will



change agriculture, but we think we are onto a real opportunity to improve photosynthesis.

What has been the biggest challenge during your first year in this position?

The hardest thing about starting a lab is recruiting. I'm competing with high-profile researchers who have established track records and funding. Yet it's crucial to get good people. No matter how good a scientist you are, you only have so much time. I've hired five people so far.

How did you overcome that challenge?

The key is to be active. I e-mail friends and colleagues and let them know I'm looking for excellent people. I can offer exciting projects not being done in other labs. I've also made it clear that I will help new hires to get what they need to make their dreams come true. Many will want faculty positions after leaving my lab. To get them, they might need something I can't provide, such as a letter from someone established in the field, but I will help them to get those letters.

So you could be competing against your postdocs for jobs?

We've created a niche for ourselves in the field of functional genomics of plants, so hopefully there will be plenty of room for us all to have exciting careers. I am laying plans to avoid competition and create win-win situations for everybody.

Past attempts to improve photosynthesis have failed. Are you concerned?

Yes. It is risky and we may not achieve it. But given our approach, we're bound to discover important and fascinating biology. ■

INTERVIEW BY VIRGINIA GEWIN

AGRICULTURAL SCIENCE

African spending up

Research opportunities have emerged in some sub-Saharan nations as a result of their increased agricultural-research spending between 2000 and 2008. *African Agricultural R&D in the New Millennium: Progress for Some, Challenges for Many*, released on 7 April by the International Food Policy Research Institute in Washington DC, surveys 32 countries. The region's total agricultural-research budget was US\$1.7 billion in 2008, up from \$1.4 billion in 2001; Nigeria alone contributed some 23% of the latest figure. Nigeria and other countries have increased salary levels and improved infrastructure, which has resulted in more researchers being hired, says report co-author Nienke Beintema. Ghana, Tanzania and Uganda showed similar trends. But spending fell in nations such as Ethiopia and South Africa.

CLEAN ENERGY

Boost for solar research

A US federal grant to fund photovoltaic research is expected to create hundreds of academic and industrial jobs. On 5 April, the US Department of Energy awarded US\$57 million to the College of Nanoscale Science and Engineering (CNSE), part of the State University of New York at Albany, to support the Photovoltaic Manufacturing Initiative, a partnership between academia and industry that aims to help the nation regain competitiveness in solar technologies. Administrators say the grant will lead to jobs for physicists, materials scientists and chemists, as well as graduate-fellowship opportunities. Pradeep Haldar, a nanoengineer at the CNSE, expects that within five years, large manufacturers will create jobs to take advantage of the college's expertise and technology.

BIOMEDICINE

NIH spared budget slash

The US National Institutes of Health (NIH) has dodged major disruption. As part of President Barack Obama's budget deal for fiscal year 2011, the agency's funds have been cut by US\$260 million, less than 1% of its \$31-billion spending plan, rather than by the \$1.6 billion sought by Republicans. "This is a bruising rather than a big gash," says Bill Talman, president of the Federation of American Societies for Experimental Biology (FASEB) in Bethesda, Maryland. The 2012 budget will be the next fight: FASEB is advocating that the NIH increase spending to \$35 billion.

ROUNABOUTS

A coming of age.

BY SUSAN LANIGAN

“Christ,” Diana says. “I’ve never seen so many of them.” She means the roundabouts, 13 in all, designed to bypass Jake’s hometown. Twelve of them are named after the Twelve Chieftains of Macha; the thirteenth after a local councillor of dubious integrity.

Jake only laughs in reply. His profile in the seat beside her is like a blade, no soft forgiveness of chin, cheek or forehead. Watchful. They are all like that here.

Diana hadn’t told anyone where she was going, except Dr Anand, yesterday. His response was to shrug his shoulders and continue to fill the syringe with the chemical cocktail for Women’s Extender treatment. First the norethisterone, enough of it to extend Diana’s menstrual cycle to a two-month interval and slow down egg death, then the hydrogen peroxide mix, which would bind to her eggs and preserve them until they were released into the body, then a weak ammonium compound to keep the mixture stable.

Jake drives past a billboard of a woman in her forties wearing nothing but a padded bra, a spoon of diet yoghurt in her mouth. Diana sighs with relief. If they have sexy posters of women her age, it must be a friendly town. Some places haven’t been so welcoming. One village in Midwarthenshire successfully chased two Extender women out of town, ‘Trouts Out’ having been painted all over their houses.

Jake pulls up at a farmhouse. “We’re home!” he calls, leading Diana into a hall with striped fleur-de-lys wallpaper. His mother rushes up to welcome them. “Och, it’s lovely to meet you, Diana. Will ye not come in and have some tea and sandwiches?”

The tea is strong and scalding. “I hope the journey was all right, pet,” his mother smiles. “You know wee Jake, his generation always act as if cars are toys to play with.”

His generation. Diana immediately recognizes the deliberate jibe. Jake, oblivious, leaves for the bathroom. Mrs McCrea waits a moment before speaking again.

“I’m a churchgoing person, y’know. I have a lot of friends on the other side who feel the same way as I do about things.” Diana just nods her head, not mentioning how Jake’s fingers whiten when he drives through areas with the wrong colour flags flying. “Both the priest on the other side and I oppose those women freaks. It’s disgusting, you know.



Having babies when they’re ready for their pension.” She reaches out to pat Diana’s arm, but Diana recoils.

“It’s not disgusting. Society has changed. They test people regularly to make sure they’re fit for purpose.”

“Och yes,” Mrs McCrea says softly. “They keep telling lies about not getting defective wee babies any more. But you and I know those eggs are full of chemicals so they’ll keep and not go off. Full of poison, so they are.” Mrs McCrea leans close. “But it’s not about testing, or society, I’m talking about. Don’t ye think it would be better for Jake to be with a nice young gerrul?”

Before Diana can exclaim with anger, Mrs McCrea imperiously lifts her hand. “I knew the minute I saw you. Could smell the hormones a mile off. D’ye know, someone from this town tried that once? The puir auld lady, sure you know what is in them syringes? Hydrogen peroxide. Bleach. Well that’s what happened. They forgot to dilute it. Bleach all over, burning her flesh from inside. She died, you know. Died roaring —”

Diana gets up and walks out without another word. Behind her the witch is laughing. Laughing! Diana

puts her hands over her ears to block it out.

“Bleach” — that is one of the many rumours about the Extender treatment. Infertility, endometriosis, unspeakable cancers. Even a recent High Court case. And now she can feel the poison coursing through her, just as Mrs McCrea threatened.

Jake is here now, talking at her through a fog: “It’s just Ma’s way. You know how sometimes we tell stories — it’s our way of saying things rather than directly.”

“Really Jake? Is that why you’ve got so many bloody roundabouts? Because you people can’t do anything straight?”

Jake starts shouting back. He is angry at her words. Her English pride, he calls it. But Diana cannot answer him. Her gut is heaving, her face pale, forehead covered in sweat. She runs away toward the barn and pukes a streaky brown mixture onto the ground. Like Mrs McCrea’s tea.

The following morning she is far away, back in London with Dr Anand. He is so solicitous, so gentle, that she breaks down in his arms. He has probably heard stories like this from many Extender women. But still he says no.

“What do you mean, no? I want to end it. Are you trying to kill me?”

And then he starts laughing, taking off his glasses and wiping the bridge of his nose, air hissing through his teeth.

“Ah Diana! No, no. *That’s* not why you are ill.”

Jake, when finally reached on his mobile, is incredulous. “You’re not, are you? You’re *not*!”

“Two months.”

“Two months! But — that’s fantastic, Di. I’ll come over.”

“Jake —”

“Love you, babe. Gotta go now.”

He rings off. Diana sips some ginger tea and swallows down another wave of nausea. She has just decided. It is not going to happen.

Dr Anand will be disappointed; she is a poster girl for Extender pregnancy. So will Jake. But she cannot go through with this. Night after night she dreams of Mrs McCrea and wakes screaming. Bleach. Poison. The smell of hormones.

She will just have to find some way, at some stage, of telling Jake the truth — without actually telling him. ■

Susan Lanigan is a programmer and writer. She lives on the east coast of Ireland, near Dublin city.

ABOUT THE JOURNAL SCOPE

Clinical Pharmacology & Therapeutics is the authoritative, cross-disciplinary journal in experimental and clinical medicine devoted to publishing advances in the nature, action, efficacy and evaluation of therapeutics in humans. *CPT* welcomes original reports and articles in the emerging areas of translational, predictive and personalized medicine; new therapeutic modalities including gene and cell therapies; pharmacogenomics, proteomics and metabolomics; bioinformation and applied systems biology complementing areas of pharmacokinetics and pharmacodynamics, human investigation and clinical trials, pharmacovigilance, pharmacoepidemiology, pharmacometrics, and population pharmacology. The journal also publishes articles by invitation. Commentary and Point-Counterpoint provide a forum for perspectives in clinical pharmacology and therapeutics in the context of contemporary scientific, political, economic and social issues. State of the Art contributions summarize the latest advances in the science underpinning drug discovery, development, regulation and utilization. *CPT* highlights issues transforming the practice of clinical pharmacology, including ethics, education and public policy. Bench-to-bedside translation in therapeutics is presented in the context of clinical applications of basic pharmacology, discovery and translational medicine and drug development. *CPT* will not consider case reports or animal studies.

ABSTRACTING

This journal is covered by: Biological Abstracts, BIOSIS Previews, Chemical Abstracts, Current Contents, Derwent Biotechnology Abstracts, EMBASE, Index to Scientific Reviews, Inpharma Weekly, International Pharmaceutical Abstracts, Medline, Nutrition Abstracts, Nutrition Research Newsletter, Personal Alert, Pharmacoeconomics and Outcomes News, Protozoological Abstracts, Reactions of Aromatic and Medicinal Plants, Review of Medical and Veterinary Mycology, Science Citation, Scopus and Tropical Diseases Bulletin.

IMPACT FACTOR

With an impact factor of 6.961, *CPT* is the top-ranked journal publishing primary research in the field of pharmacology and pharmacy.

URL <http://www.nature.com/cpt>

ISSN 0009-9236

EISSN 1532-6535

FREQUENCY Printed monthly. Advance Online Publication once weekly.

PREPARATION OF MANUSCRIPTS CONTENT TYPES

Author Submissions (1) Article, (2) Clinical Trial, (3) Letter to the Editor, (4) Report

Invited Submissions (5) Book Review, (6) Commentary, (7) Conference Proceedings, (8) Development, (9) Discovery, (10) Macroscopy, (11) Opinions, (12) Point-Counterpoint, (13) Practice, (14) Review, (15) State of the Art, (16) Translation. Submissions that do not adhere to the guidelines provided in this document will be returned to the author prior to consideration. Material that cannot fit within the allowed limit may be submitted as supplementary information.

ORIGINAL RESEARCH

(1) Article

Word Limit: 4,000 words excluding abstract, references, tables, and figures

Abstract: 150 words maximum

References: 50 maximum

Figures/Tables: 7 maximum

Substantial novel research

(2) Clinical Trial

Word Limit: 5,000 words excluding abstract, references, tables, and figures

Abstract: 150 words maximum

References: 60 maximum

Figures/Tables: 8 maximum

Manuscripts developed from well-conducted, well reported, and relevant clinical trials. See Clinical Trials: Guide to Authors for additional information.

(3) Letter to the Editor

Word Limit: 400 words excluding references, tables, and figures

Abstract: no abstract for this article type

References: 5 maximum

Figures/Tables: 1 maximum

Typically should address issues concerning recently published information in Clinical Pharmacology & Therapeutics. A Letter to the Editor must reference the original source, and a Response to Letter must reference the Letter to the Editor in the first few paragraphs. Letters to the Editor can use an arbitrary title, but a Response must cite the title of the Letter: e.g., Response to [title of Letter]

(4) Report

Word Limit: 2,000 words excluding abstract, references, tables, and figures

Abstract: 75 words maximum

References: 30 maximum

Figures/Tables: 3 maximum

Substantial novel research

OTHER CONTRIBUTIONS

(5) Book Review [only by invitation of Editors]

Word Limit: 700 words

Abstract: no abstract for this article type

References: no references for this article type

Figures/Tables: Cover image will be secured prior to publication by the Editorial Office

(6) Commentary [only by invitation of Editors]

Word Limit: 1,600 words excluding abstract, references, tables, and figures

Abstract: no abstract for this article type; should include a 75-word introduction

References: 10 maximum

Figures/Tables: 2 maximum

Typically highlights findings of a paper in the same issue, presented in a wider scientific and clinical context

(7) Conference Proceedings [only by invitation of Editors]

Word Limit: 1,600 words excluding abstract, references, tables, and figures

Abstract: no abstract for this article type; should include a 75-word introduction

References: 10 maximum

Figures/Tables: 2 maximum

Typically highlights topics from recent conferences that impact clinical pharmacology

(8) Development [only by invitation of Editors]

Word Limit: 3,000 words excluding references, tables, and figures

Abstract: no abstract for this article type; should include a 75-word introduction

References: 40 maximum

Figures/Tables: 3 maximum

Illustrations: 1 required

Emerging innovations in tools, platforms, and applications transforming diagnostic and therapeutic development; new algorithms optimizing the clinical utility of patient- and

population-based interventions; evolution of regulatory and economic policies centered on optimizing disease management

(9) Discovery [only by invitation of Editors]

Word Limit: 3,000 words excluding references, tables, and figures

Abstract: no abstract for this article type; should include a 75-word introduction

References: 40 maximum

Figures/Tables: 3 maximum

Illustrations: 1 required

Cellular or molecular mechanisms emerging from enabling platform technologies providing insight into pathophysiology, opening new avenues for diagnostic and therapeutic intervention; novel integration of fundamental principles across communities of practice and disciplines producing unexpected paradigms with transformative clinical potential

(10) Macroscopy [only by invitation of Editors]

Word Limit: 1,600 words

Abstract: no abstract for this article type

References: 5 maximum

Figures/Tables: 1 maximum

Should offer a broad view on critical issues facing clinical pharmacology and therapeutics in science, healthcare, policy, and society

(11) Opinions

Word Limit: 1,500 words excluding references, tables, and figures

Abstract: no abstract for this article type; should include a 75-word introduction

References: 5 maximum

Figures/Tables: 2 maximum

These short pieces are designed to give the author's perspective on current topics of relevance to the readership in areas of education, ethics, and public policy

(12) Point-Counterpoint [only by invitation of Editors]

Word Limit: 1,600 words excluding abstract, references, tables, and figures

Abstract: no abstract for this article type; should include a 75-word introduction

References: 10 maximum

Figures/Tables: 2 maximum

Balanced discussion of controversies in clinical pharmacology

(13) Practice [only by invitation of Editors]

Word Limit: 2,000 words excluding references, tables, and figures.

Abstract: no abstract for this article type; should include a 75-word introduction

References: 30 maximum

Figures/Tables: 3 maximum

Illustrations: 1 required

Cases of exceptional novelty that hold teaching points applicable to clinical pharmacology and established therapeutic approaches in clinical care

(14) Review [only by invitation of Editors]

Word Limit: 8,000 words excluding references, tables, and figures

Abstract: no abstract for this article type; should include a 75-word introduction

References: 75 maximum

Figures/Tables: 8 maximum

Illustrations: 2 required

High-quality, timely reviews and perspectives covering important topics in the entire field of clinical pharmacology and therapeutics

(15) State of the Art [only by invitation of Editors]

Word Limit: 8,000 words excluding references, tables, and figures

Abstract: 150 words maximum

References: 75 maximum

Figures/Tables: 8 maximum

Illustrations: 2 required

Typically topical reviews, award lectures, keynote addresses, and State of the Art Lectures

(16) Translation [only by invitation of Editors]

Word Limit: 3,000 words excluding references, tables, and figures

Abstract: no abstract for this article type; should include a 75-word introduction

References: 40 maximum

Figures/Tables: 3 maximum

Illustrations: 1 required

Emerging innovations in tools, platforms, and applications transforming diagnostic and therapeutic development; new algorithms optimizing the clinical utility of patient- and population-based interventions; evolution of regulatory and economic policies centered on optimizing disease management

FORMAT OF MANUSCRIPTS

General format Manuscripts must be typed in English and be in a single column, double-spaced format. All manuscript pages must be numbered.

Title page This should include (a) the complete manuscript title; (b) all authors' names (listed as first and middle initials followed by last name) and affiliations; and (c) the name and address for correspondence, fax number, telephone number, and e-mail address. The title page should also include word counts for the abstract, introduction, and the manuscript as a whole; the number of references, figures, and tables; and key words.

Text For contributions requiring abstracts, the lengths are defined in the respective sections of *Preparation of Manuscripts*. For contributions that do not require an abstract, introductory paragraphs may contain references to cited work. Articles and Reports should consist of the following ordered sections:

Title Page
Abstract
Introduction
Results
Discussion
Methods (must contain IRB or IACUC approval: see **Informed Consent and Ethics** below)
Acknowledgements
Conflict of Interest/Disclosure
References
Figure Legends
Tables

Originality A submitted manuscript must be an original contribution not previously published (except as an abstract), must not be under consideration for publication elsewhere, and, if accepted, must not be reproduced elsewhere without the consent of the American Society for Clinical Pharmacology and Therapeutics (ASCP). Although the editors, editorial board, and referees make every effort to ensure the validity of published manuscripts, the final responsibility rests with the authors, not with *Clinical Pharmacology & Therapeutics*, its editors, the ASCPT, or Nature Publishing Group.

Informed Consent and Ethics CPT adheres to the *Uniform Requirements for Manuscripts Submitted to Biomedical Journals* established by the International Committee of Medical Journal Editors. A full description of recommendations can be found at <http://www.icmje.org>. Research projects involving human subjects require review and approval by an Institutional Review Board (IRB). When reporting experiments on human subjects, indicate whether the procedures were in accordance with the ethical standards of the responsible committee on human experimentation or with the Helsinki Declaration of 1975 (as revised in 1983). IRB or IACUC approval must be cited in the Methods section of the text. If there has been no IRB review of the study, please so indicate in the cover letter. In such situations, the manuscript

will be reviewed to determine if IRB review should have been conducted. The result of this review may determine whether or not the manuscript will be considered for publication.

Clinical Trials Registry Registration in a public trials registry is required for publication in CPT. A clinical trial is defined as any research project that prospectively assigns human subjects to intervention or comparison groups to study the cause-and-effect relationship between a medical intervention and a health outcome. Studies designed for other purposes, including exploring pharmacokinetics or safety and tolerability (e.g., phase 1 trials) are exempt. Registration must be with a registry that meets the following criteria: (1) accessible to the public at no charge; (2) searchable by electronic methods; (3) open to all prospective registrants free of charge or at minimal cost; (4) validates registered information; (5) identifies trials with a unique number; and (6) includes information on the investigator(s), research question or hypothesis, methodology, intervention and comparisons, eligibility criteria, primary and secondary outcomes measured, date of registration, anticipated or actual start date, anticipated or actual date of last follow-up, target number of subjects, status (anticipated, ongoing or closed), and funding source(s). Examples of registries that meet these criteria include (1) The registry sponsored by the United States National Library of Medicine (<http://www.clinicaltrials.gov>); (2) The International Standard Randomised Controlled Trial Number Registry (<http://www.controlled-trials.com>); (3) The Cochrane Renal Group Registry (<http://www.cochrane-renal.org/trials/submitform.php>); (4) The National (United Kingdom) Research Register (<http://www.update-software.com/national/>); and (5) European Clinical Trials Database (<http://eudract.emea.eu.int/>).

Abbreviations Abbreviations should be defined at the first mention in the text and in each table and figure. For a list of standard abbreviations, please consult the CSE Manual for Authors, Editors, and Publishers (available from the Council of Science Editors, 12100 Sunset Hills Road, Suite 130, Reston, VA 20190) or other standard sources. Write out the full term for each abbreviation at its first use unless it is a standard unit of measure.

Style *The American Medical Association Manual of Style* (9th edition), *Stedman's Medical Dictionary* (27th edition) and *Merriam-Webster's Collegiate Dictionary* (10th edition) should be used as standard references. Refer to drugs and therapeutic agents by their accepted generic or chemical name, and do not abbreviate them (a proprietary name may be given only with the first use of the generic name). Code names should be used only when a generic name is not yet available (the chemical name and a figure giving the chemical structure of the drug is required). Copyright or trade names of drugs should be capitalized and placed in parentheses after the name of the drug. Names and locations (city and state in United States; city and country outside United States) of manufacturers of drugs, supplies, or equipment cited in a manuscript are required to comply with trademark law and should be provided in parentheses.

Language Editing Authors who require editing for language are encouraged to consult language editing services prior to submission. Contact the Editorial Office for recommendations.

AUTHOR RESPONSIBILITY

Upon submission, the corresponding author must confirm full access to all data in the study and final responsibility.

ACKNOWLEDGMENTS

This should include sources of support, including federal and industry support. All authors who have

contributed to the manuscript must be acknowledged. Medical writers, proofreaders, and editors should not be listed as authors, but acknowledged at the beginning or end of the text.

DISCLOSURE

At the time of submission, each author must disclose and describe any involvement, financial or otherwise, that might potentially bias his or her work. Disclosure must be included in the text of the manuscript.

REFERENCES

Should be listed in order of appearance (Vancouver style). In the text, number references in order of appearance using Arabic numerals (e.g., 1, 2, 3) in parentheses for citations. The publisher will convert parenthetical citations to superscript at the proofing stage. The reference list (starting on a separate page) should contain the references in the order in which they are cited in the text. Citations included in tables/figures count toward the maximum references allowed for the article type and must be included in the reference list. Tables created solely of references are not permitted. Only published works, as well as manuscripts in press, should be included in the reference list; articles that are submitted or in preparation should be referred to as "unpublished data" in the text (for which all authors up to 6 total should be listed, then et al.). For publications in the reference list, all authors should be included unless there are more than six, in which case only the first author should be given, followed by 'et al.' Authors should be listed last name first, followed by a comma and initials of given names. Titles of cited articles are required for all article types. Titles of articles should be in Roman text and titles of books in italics. The titles should be written exactly as they appear in the work cited except that article titles should have only the first word capitalized, and they should end with a period. Journal names are italicized and abbreviated (with periods after each abbreviated word) according to common usage; refer to Index Medicus (PubMed) for details. Volume numbers appear in bold. For book citations, the publisher and city of publication are required; include the country (and state for US) for lesser-known cities or where any ambiguity is possible (e.g., John Wiley & Sons, Hoboken, New Jersey, USA, 2003; MIT Press, Cambridge, Massachusetts, USA). Please note the following examples:

Journal articles:

Kashuba, A.D. et al. Effect of fluvoxamine therapy on the activities of CYP1A2, CYP2D6, and CYP3A as determined by phenotyping. *Clin. Pharmacol. Ther.* **64**, 257–268 (1998).

Books:

Eisen, H.N. *Immunology: An Introduction to Molecular and Cellular Principles of the Immune Response* 5th edn. (Harper & Row, New York, 1974).

Articles in books:

Weinstein, L. & Schwartz, M.N. Pathogenic properties of invading microorganisms. In *Pathologic Physiology: Mechanisms of Disease* (eds. Sodeman, W.A., Jr. & Sodeman, W.A.) 457–473 (W.B. Saunders, Philadelphia, 1974).

CPT is supported by EndNote Styles. To download the CPT style file, visit <http://www.endnote.com/support/enstyles.asp> and search for "Clinical Pharmacology & Therapeutics."

FIGURES

Should be labeled sequentially, numbered, and cited in the text. If a table, figure or any other previously published material is included, the authors must obtain written permission to reproduce the material in both print and electronic formats from the copyright owner and submit it with the manuscript. The original

source should be cited. Figures and tables must be uploaded separately from the manuscript text.

FIGURE LEGENDS

Legends should be brief and specific and should appear on a separate manuscript page after the Reference section.

GUIDELINES FOR FIGURES AND ARTWORK

Detailed guidelines for submitting figures and artwork can be found at: <http://www.nature.com/aj/artworkguidelines.pdf>. Using the guidelines, please submit production quality artwork with your initial online submission. If you have followed the guidelines, we will not require unchanged artwork to be resubmitted following the peer-review process.

FIGURES IN PRINT

Accepted figure files include JPEG, TIFF, EPS, AI, and PSD.

[color charges may apply]

Minimum resolutions:

Halftone images, 300 dpi (dots per inch)

Color images, 300 dpi saved as CMYK

Images containing text, 400 dpi

Line art, 1,000 dpi

Sizes:

Figure width – single image 86 mm (should be able to fit into a single column of the printed journal)

Figure width – multi-part image 178 mm (should be able to fit into a double column of the printed journal)

Text size

8 point (should be readable after reduction – avoid large type or thick lines) line width between 0.5 and 1 point

COLOR ON THE WEB

For FREE color figures on the web (only available in the HTML (full text) version of manuscripts), authors should supply separate JPEG or GIF files. These files should be submitted as supplementary information and indicated as such in the submission letter.

For single images:

Width 500 pixels (authors should select “constrain proportions,” or equivalent instructions, to allow the application to set the correct proportions automatically)

Resolution 125 dpi (dots per inch)

Format JPEG for photographs, GIF for line drawings or charts

File naming Please save image with .jpg or .gif extension to ensure it can be read by all platforms and graphics packages

For multi-part images:

Width 900 pixels (authors should select “constrain proportions,” or equivalent instructions, to allow the application to set the correct proportions automatically)

Resolution 125 dpi (dots per inch)

Format JPEG for photographs, GIF for line drawings or charts

File naming Please save image with .jpg or .gif extension to ensure it can be read by all platforms and graphics packages

TABLES

Accepted file types include MS Word (tables should be editable, not embedded images) and Excel. Each table should be double-spaced on a separate sheet and numbered consecutively in the order of first citation in the text. Minimize empty space and restrict the number of characters per row to 130. Supply a brief title for each, but place explanatory matter in the footnotes (not in the heading). Do not use internal horizontal and vertical lines.

JOURNAL STYLE

Papers should be prepared as follows:

1. See the artwork guidelines above
2. Do not make rules thinner than 1 pt (0.36mm)
3. Use a coarse hatching pattern rather than shading for tints in graphs
4. Color should be distinct when used as an identifying tool
5. Commas should be used to separate thousands
6. Abbreviations should be preceded by the words for which they stand in the first instance of use
7. Text should be double-spaced with a 1 inch margin
8. At first mention of a manufacturer, the town (state if USA) and country should be provided

FILE FORMATS & REQUIREMENTS

File formats are provided in the online forms. Use a common word-processing package (MS Word is preferred) for the text. Please note: A Word 2007 document must be saved as a copy fully compatible with Word 97-2003 prior to submission. Tables should be provided at the end of the Word document.

Authors are required to submit final, publication-ready files at the revision stage, along with a version tracking all changes to the paper. Use the Track Changes mode in MS Word or indicate the revised text with bold, highlighted, or colored type.

SUPPLEMENTARY INFORMATION

Supplementary information is peer-reviewed material directly relevant to the conclusion of an article that cannot be included in the printed version owing to space or format constraints. It is posted on the journal's web site and linked to the article when the article is published; it may include data files, graphics, movies, or extensive tables. The printed article must be complete and self-explanatory without the supplementary information. Supplementary information must be supplied to the editorial office in its final form for peer review. On acceptance, the final version of the peer-reviewed supplementary information should be submitted with the accepted paper. To ensure that the contents of the supplementary information files can be viewed by the editor(s), referees, and readers, please also submit a ‘read-me’ file containing brief instructions on how to use the file.

Supplementary Information Charges Supplementary information may be included online at a rate of \$125 per file.

Supplying Supplementary Information Files Authors should ensure that supplementary information is supplied in its FINAL format, as it is not copy edited and will appear online exactly as submitted. It cannot be altered, nor new supplementary information added, after the paper has been accepted for publication. Please supply the supplementary information via the electronic manuscript submission and tracking system in an acceptable file format (see below). Authors should include a text summary (no more than 50 words) to describe the contents of each file, identify the types of files (file formats) submitted and include the text ‘Supplementary information is available at <http://www.nature.com/cpt>’ at the end of the body of text and before the references.

Accepted File Formats Quick Time files (.mov), graphical image files (.gif), HTML files (.html), MPEG movie files (.mpg), JPEG image files (.jpg), sound files (.wav), plain ASCII text (.txt), Acrobat files (.pdf), MS Word documents (.doc), Postscript files (.ps), MS Excel spreadsheet documents (.xls), and Powerpoint (.ppt). We cannot accept TeX or LaTeX. File sizes must be as small as possible to expedite downloading. Images should not exceed 640 x 480 pixels. For movies, we recommend 480 x 360 pixels as the maximum frame size and a frame rate of 15 frames per second. If applicable to the presentation of the information, use a 256-color palette. Please consider the use

of lower specification for all of these points if the supplementary information can still be represented clearly. Our recommended maximum data rate is 150 KB/s. The number of files should be limited to eight, and the total file size should not exceed 8 MB. Individual files should not exceed 1 MB. Seek advice from the editorial office before sending files larger than our maximum size to avoid delays in publication. Questions about the submission or preparation of supplementary information should be directed to the editorial office.

EDITORIAL POLICIES

PLAGIARISM

Plagiarism is when an author attempts to represent someone else's work as his or her own. Duplicate publication, sometimes called self-plagiarism, occurs when an author reuses substantial parts of his or her own published work without providing the appropriate references. This can range from getting an identical paper published in multiple journals, to ‘salami-slicing’, where authors add small amounts of new data to a previous paper. Plagiarism can be said to have clearly occurred when large chunks of text have been cut-and-pasted. Such manuscripts would not be considered for publication in a Nature journal. But minor plagiarism without dishonest intent is relatively frequent, for example, when an author reuses parts of an introduction from an earlier paper. The journal editors judge any case of which they become aware (either by their own knowledge of and reading about the literature, or when alerted by referees) on its own merits. If a case of plagiarism comes to light after a paper is published, the journal will conduct a preliminary investigation. If plagiarism is found, the journal will contact the author's institute and funding agencies. A determination of misconduct will lead the journal to run a statement, bidirectionally linked online to and from the original paper, to note the plagiarism and to provide a reference to the plagiarized material. The paper containing the plagiarism will also be obviously marked on each page of the PDF. Depending on the extent of the plagiarism, the paper may also be formally retracted.

IMAGE INTEGRITY AND STANDARDS

Images submitted with a manuscript for review should be minimally processed (for instance, to add arrows to a micrograph). Authors should retain their unprocessed data and metadata files, as editors may request them to aid in manuscript evaluation. If unprocessed data are unavailable, manuscript evaluation may be stalled until the issue is resolved. A certain degree of image processing is acceptable for publication (and for some experiments, fields and techniques is unavoidable), but the final image must correctly represent the original data and conform to community standards. The guidelines below will aid in accurate data presentation at the image processing level; authors must also take care to exercise prudence during data acquisition, where misrepresentation must equally be avoided. Authors should list all image acquisition tools and image processing software packages used. Authors should document key image gathering settings and processing manipulations in the Methods. Images gathered at different times or from different locations should not be combined into a single image, unless it is stated that the resultant image is a product of time-averaged data or a time-lapse sequence. If juxtaposing images is essential, the borders should be clearly demarcated in the figure and described in the legend. The use of touch-up tools, such as cloning and healing tools in Photoshop, or any feature that deliberately obscures manipulations, is to be avoided. Processing (such as changing brightness and contrast) is appropriate only when it is applied equally across the entire image and is applied equally to controls. Contrast should not be adjusted so that data disappear. Excessive manipulations, such as processing to emphasize one region in the image at

the expense of others (for example, through the use of a biased choice of threshold settings), is inappropriate, as is emphasizing experimental data relative to the control. When submitting revised final figures upon conditional acceptance, authors may be asked to submit original, unprocessed images.

CONFIDENTIALITY

CPT editors and editorial staff keep confidential all details about a submitted manuscript and do not comment to any outside organization about manuscripts under consideration by the journal while they are under consideration or if they are rejected. The journal editors may comment publicly on published material, but their comments are restricted to the content itself and their evaluation of it. After a manuscript is submitted, correspondence with the journal, referees' reports and other confidential material, whether or not the submission is eventually published, must not be posted on any website or otherwise publicized without prior permission from the editors. The editors themselves are not allowed to discuss manuscripts with third parties or to reveal information about correspondence and other interactions with authors and referees. Referees of manuscripts submitted to CPT undertake in advance to maintain confidentiality of manuscripts and any associated supplementary data.

PRE-PUBLICITY

Authors must not discuss contributions with the media (including other scientific journals) until the publication date; advertising the contents of any contribution to the media may lead to rejection. The only exception is in the week before publication, during which contributions may be discussed with the media if authors and their representatives (institutions, funders) clearly indicate to journalists that their contents must not be publicized until the journal's press embargo has elapsed. Authors will be informed of embargo dates and timings after acceptance for publication of their articles.

COMMUNICATION WITH THE MEDIA

Material submitted to CPT must not be discussed with the media, except in the case of accepted contributions, which can be discussed with the media no more than a week before the publication date under our embargo conditions. We reserve the right to halt the consideration or publication of a paper if this condition is broken. From time to time CPT will distribute to a registered list a press release summarizing selected content of the next issue's publication. Journalists are encouraged to read the full version of any papers they wish to cover, and are given the names of corresponding authors, together with phone and fax numbers and email addresses. They receive access to the full text of papers about a week before publication on a password-protected website, together with other relevant material (for example, an accompanying News and Views article, and any extra illustrations provided by the authors). The content of the press release and papers is embargoed until the time and date clearly stated on the press release. Authors may therefore receive calls or emails from the media during this time; we encourage them to cooperate with journalists so that media coverage of their work is accurate and balanced. Authors whose papers are scheduled for publication may also arrange their own publicity (for instance through their institutional press offices), but they must strictly adhere to our press embargo.

COMMUNICATION BETWEEN SCIENTISTS

CPT does not wish to hinder communication between scientists. For that reason, different embargo guidelines apply to work that has been discussed at a conference or displayed on a preprint server and picked up by the media as a result. (Neither conference presentations nor posting on recognized preprint servers constitute prior publication.)

CORRECTION AND RETRACTION POLICY

We recognize our responsibility to correct errors that we have previously published. Our policy is to consider refutations (readers' criticisms) of primary research papers, and to publish them (in concise form) if and only if the author provides compelling evidence that a major claim of the original paper was incorrect. Corrections are published for significant errors in non-peer-reviewed content of the Nature journals at the discretion of the editors. Readers who have identified such an error should send an email to the editorial office of the journal, clearly stating the publication reference, title, author and section of the article, and briefly explaining the error.

CORRECTIONS TO THE PRINT AND ONLINE VERSIONS OF PEER-REVIEWED CONTENT

Publishable amendments requested by the authors of the publication are represented by a formal printed and online notice in the journal because they affect the publication record and/or the scientific accuracy of published information. Where these amendments concern peer-reviewed material, they fall into one of four categories: erratum, corrigendum, retraction, or addendum, described here.

Erratum Notification of an important error made by the journal that affects the publication record or the scientific integrity of the paper, or the reputation of the authors, or of the journal.

Corrigendum Notification of an important error made by the author(s) that affects the publication record or the scientific integrity of the paper, or the reputation of the authors or the journal. All authors must sign corrigenda submitted for publication. In cases where coauthors disagree, the editors will take advice from independent peer-reviewers and impose the appropriate amendment, noting the dissenting author(s) in the text of the published version.

Retraction Notification of invalid results. All coauthors must sign a retraction specifying the error and stating briefly how the conclusions are affected, and submit it for publication. In cases where coauthors disagree, the editors will seek advice from independent peer reviewers and impose the type of amendment that seems most appropriate, noting the dissenting author(s) in the text of the published version.

Addendum Notification of a peer-reviewed addition of information to a paper, usually in response to readers' request for clarification. Addenda are published only rarely and only when the editors decide that the addendum is crucial to the reader's understanding of a significant part of the published contribution.

SUBMISSION AND PUBLICATION

SUBMISSION OF PAPERS

Manuscripts must be submitted online at <http://mts-cpt.nature.com>. Manuscripts are assessed by an editor upon submission. Only manuscripts that meet our editorial criteria are sent out for formal review. One compelling, negative review may be sufficient for a decision to reject.

Submission Fee (Does not apply to invited authors) Manuscripts submitted for consideration will incur a submission fee of \$75 to cover, in part, the time and resources required to manage submissions. Members of ASCPT will receive a discounted submission rate. Fees must be paid prior to final submission of a paper and will not be waived or refunded.

COPYRIGHT

Ownership is to be transferred to the American Society for Clinical Pharmacology and Therapeutics. The Copyright Statement form must be signed and returned to the editorial office prior to publication. Failure to do so will result in delays to the publication of your paper. A single designated author may sign the

Copyright Statement form on behalf of all authors. The enclosure of a copyright transfer form in a request for a revised manuscript does not imply that the revised manuscript will be accepted. The Copyright Statement is also available under "Instructions & Forms" on the online submission page: <http://mts-cpt.nature.com>.

NIH Open Access Policy In compliance with the National Institutes of Health Open Access Policy, ASCPT grants limited copyright release to authors who have received funding for research on which the article is based from the National Institutes of Health to deposit the accepted paper into PubMed Central. It remains the legal responsibility of the author(s) to deposit the final, peer-reviewed paper upon acceptance for publication, to be made publicly available no sooner than 12 months after the official date of publication.

PAGE AND COLOR CHARGES

(Do not apply to invited authors)

Page charges Manuscripts accepted for publication in Clinical Pharmacology & Therapeutics will incur page charges to cover, in part, the cost of publication. A charge of \$50 will be issued for each journal page.

Color charges Authors will be expected to contribute towards the cost of publication of color figures. A quote will be supplied upon acceptance of the paper.

Charges are:

- 1 figure: \$846
- 2 figures: \$1,260
- 3 figures: \$1,674
- 4 figures: \$1,926
- 5 figures: \$2,178
- 6 figures: \$2,394
- 7+ figures: \$216 per additional figure

Upon acceptance authors must fill in the color artwork form available at <http://mts-cpt.nature.com>.

Offprints May be ordered using the order form available for download with the proofs.

PROOFS

An email will be sent to the corresponding author with a URL link from where proofs can be collected. Proofs must be returned by fax within 48 hours of receipt. Failure to do so may result in a delay to publication. Extensive corrections cannot be made at this stage.

ADVANCE ONLINE PUBLICATION

All original articles and reviews are published ahead of print on Advance Online Publication. This will be the final version of the manuscript and will subsequently appear, unchanged, in print.

CONTACT INFORMATION

EDITORIAL

All business regarding manuscripts and peer review should be addressed to:

Clinical Pharmacology & Therapeutics
528 North Washington Street
Alexandria, VA 22314

USA

Tel: +1.703.836.6981

Fax: +1.703.836.6996

Attn: cpt@ascpt.org, Managing Editor & Director of Publications
cpt2@ascpt.org, Senior Editorial Coordinator

BUSINESS MATTERS

All business correspondence and inquiries should be addressed to:

Clinical Pharmacology & Therapeutics
Nature Publishing Group
75 Varick Street, 9th Floor
New York, NY 10013
USA

Tel: +1.212.726.9301

Learning-related feedforward inhibitory connectivity growth required for memory precision

Sarah Ruediger^{1*}, Claudia Vittori^{1,2*}, Ewa Bednarek¹, Christel Genoud¹, Piergiorgio Strata², Benedetto Sacchetti² & Pico Caroni¹

In the adult brain, new synapses are formed and pre-existing ones are lost, but the function of this structural plasticity has remained unclear^{1–5}. Learning of new skills is correlated with formation of new synapses^{6–8}. These may directly encode new memories, but they may also have more general roles in memory encoding and retrieval processes². Here we investigated how mossy fibre terminal complexes at the entry of hippocampal and cerebellar circuits rearrange upon learning in mice, and what is the functional role of the rearrangements. We show that one-trial and incremental learning lead to robust, circuit-specific, long-lasting and reversible increases in the numbers of filopodial synapses onto fast-spiking interneurons that trigger feedforward inhibition. The increase in feedforward inhibition connectivity involved a majority of the pre-synaptic terminals, restricted the numbers of c-Fos-expressing postsynaptic neurons at memory retrieval, and correlated temporally with the quality of the memory. We then show that for contextual fear conditioning and Morris water maze learning, increased feedforward inhibition connectivity by hippocampal mossy fibres has a critical role for the precision of the memory and the learned behaviour. In the absence of mossy fibre long-term potentiation in *Rab3a*^{−/−} mice⁹, c-Fos ensemble reorganization and feedforward inhibition growth were both absent in CA3 upon learning, and the memory was imprecise. By contrast, in the absence of adducin 2 (*Add2*; also known as β -adducin)¹⁰ c-Fos reorganization was normal, but feedforward inhibition growth was abolished. In parallel, c-Fos ensembles in CA3 were greatly enlarged, and the memory was imprecise. Feedforward inhibition growth and memory precision were both rescued by re-expression of *Add2* specifically in hippocampal mossy fibres. These results establish a causal relationship between learning-related increases in the numbers of defined synapses and the precision of learning and memory in the adult. The results further relate plasticity and feedforward inhibition growth at hippocampal mossy fibres to the precision of hippocampus-dependent memories.

To determine whether hippocampus-dependent learning^{11–13} may produce structural rearrangements in hippocampal large mossy fibre terminal (LMT) components involved in feedforward excitation and/or feedforward inhibition in CA3 (ref. 14) (Fig. 1a and Supplementary Material), we analysed GFP-positive LMTs in the dorsal hippocampus of *Thy1-mGFP(Lsi1)* reporter mice⁵ that had been subjected to contextual fear conditioning, a one-trial learning protocol (Methods). Fear conditioning led to a robust increase in the average number of filopodia per LMT (1.82-fold, $P < 0.001$; feedforward inhibition connectivity; Fig. 1b, c and Supplementary Fig. 2a), and to a less pronounced increase in the average numbers of Bassoon-positive putative release sites per core LMT¹⁵ (1.31-fold, $P < 0.01$; feedforward excitation connectivity; Supplementary Fig. 2a). By contrast, there was no change in the densities of LMTs in CA3b at any time upon fear conditioning (Supplementary Fig. 2a). The filopodia contacted spine-free dendrites of parvalbumin-positive interneurons in CA3b (Fig. 1d, e and Supplementary Fig. 3a), indicating that they induce feedforward inhibition through fast-spiking

interneurons^{16–18}. To estimate the fraction of LMTs in CA3b with altered contents of filopodia, we analysed LMT/filopodia distributions in naive, control and fear-conditioned mice. Shifts in the fractions of LMTs with no filopodia and with more than four filopodia revealed that, on average, at least 45% of the LMTs established increased numbers of filopodia as a consequence of fear conditioning (Fig. 1f).

To determine whether an increase in stratum lucidum feedforward inhibition connectivity may be generally associated with hippocampal learning, we analysed mice that underwent a Morris water maze incremental learning protocol. Filopodial contents were only slightly increased over naive values during the first 3–4 days of training, whereas they increased markedly between days 4 and 8 (Fig. 1g). Again, we detected no changes in the densities of LMTs in CA3b upon Morris water maze learning (not shown). Testing mice for the memory of the platform position revealed that this reference memory only began to differ from chance after 3 days of training (Fig. 1h). The reference memory reached plateau values at day 8 (Fig. 1h), suggesting that filopodial growth correlated with the establishment of a precise spatial memory in the Morris water maze test. The reference memory of the platform position persisted for at least 45 days after cessation of the training and, unlike in the fear conditioning experiment, raised filopodia per LMT values also persisted for at least 45 days (Fig. 1h; day 53 values). As in the fear conditioning experiment, a large fraction of the LMTs exhibited higher filopodial contents at plateau values (Fig. 1i).

To determine whether learning-related induction of feedforward inhibition connectivity growth might be a general phenomenon not restricted to spatial learning in the hippocampus, we analysed mossy fibre terminals in the cerebellar cortex, which also consist of powerful large core structures associated with filopodia¹⁹. Cued fear conditioning, in which animals learn that a tone predicts an aversive stimulus, involves plasticity in cerebellar cortex lobule 5, but not lobule 9 (ref. 20). In parallel, cued fear conditioning led to a robust and reversible increase of filopodial numbers per mossy fibre terminal in lobule 5, but not lobule 9 (Fig. 2a, d). In a second set of experiments, we trained mice to balance on an accelerating rotating rod (rotarod). This cerebellum-dependent motor skill task involved incremental learning over 4–6 days, which was accompanied by a parallel increase in the filopodial contents of mossy fibre terminals in lobule 9, but not lobule 5 (Fig. 2b, d). At least for the Golgi cells that could be visualized with the marker RC3, mossy fibre terminal filopodia extended along their dendrites, and established numerous varicosities, where synaptic markers co-distributed (Fig. 2c and Supplementary Fig. 4). More than 95% of the filopodial varicosities within a granule cell layer volume exhibiting an RC3-positive Golgi cell made putative synaptic contacts with that Golgi cell. Therefore, learning is specifically correlated with the growth of feedforward inhibition connectivity in both hippocampal and cerebellar circuits.

We next sought to determine what might be the function of the learning-related growth in feedforward inhibition connectivity. In the fear conditioning experiments, the excess filopodia were lost within 8–10 days after learning, and filopodial retention was prolonged upon re-exposure to context leading to extinction (Fig. 3a), indicating that

¹Friedrich Miescher Institute, Maulbeerstrasse 66, CH-4058 Basel, Switzerland. ²Department of Neuroscience and National Institute of Neuroscience-Italy, C.so Raffaello 30, 10125 Torino, Italy.

*These authors contributed equally to this work.

the excess filopodia are not a requirement for expression of the fear memory. Testing of individual mice during the Morris water maze training protocol revealed a strong correlation between the reference memory of the platform position and mean filopodial contents per LMT for individual mice (Fig. 3b), indicating that the extent of filopodial growth was correlated to the precision of the learning. We therefore monitored generalization, that is, decreased behavioural precision of the fear memory in the contextual fear conditioning experiment. In agreement with previous reports^{21,22}, generalization of the memory for context in fear conditioning was not detectable during the first 5–7 days after learning, but was detected at longer intervals after fear conditioning as an enhanced freezing response and reduced exploratory activity in a neutral context (Fig. 3c). A brief re-exposure of mice to training context in the absence of the aversive stimulus at 15 days after learning produced a suppression of generalization at retest, which lasted 8–12 days (Fig. 3d). In parallel, training context re-exposure induced a pronounced re-induction of the filopodial response, which again lasted for 7–10 days (Fig. 3d). By contrast, exposure to a neutral context affected neither generalization nor filopodial growth (Fig. 3d), suggesting that retrieval of the specific memory was necessary to re-induce feedforward inhibition connectivity growth in hippocampal CA3, and to suppress generalization.

To investigate a possible functional correlate of feedforward inhibition connectivity growth, we analysed c-Fos-positive pyramidal neurons in CA3b in the contextual fear conditioning experiment²³. On day

0, mice were exposed to the training context without or with aversive stimulus. In the absence of aversive conditioning, re-exposure on day 1 to either the training context or a neutral context produced closely comparable increases in the fractions of pyramidal neurons with high and intermediate c-Fos signals when compared to naive cage control mice (Fig. 4a). In stark contrast, association of the training context with an aversive stimulus led to a specific and robust relative increase in the number of pyramidal neurons expressing high c-Fos signals upon recall of the memory in the training context, and to a marked reduction of the high and medium c-Fos signals upon exposure to the neutral context (Fig. 4a). Recall in the training context at day 15 led to decreased high-signal c-Fos neurons, whereas exposure to a neutral context at day 15 led to markedly increased low-signal c-Fos neurons (Fig. 4b). Notably, in parallel to increased filopodial numbers and the re-establishment of memory precision, memory recall in the training context at day 15 after fear conditioning suppressed excess responses upon subsequent exposure to a neutral context (Fig. 4b).

To address the role of mossy fibres and their plasticity in fear memory precision, we carried out fear conditioning experiments in *Rab3a*^{-/-} mice, which specifically lack long-term potentiation (LTP) at mossy fibres, but not at other synapses in the hippocampus⁹. We found that in the absence of *Rab3a*, mice learned the relationship between the training context and the aversive stimulus, but already generalized 1 day after fear conditioning (Fig. 4c). In parallel, *Rab3a*^{-/-} mice lacked any learning-related increase in putative release sites at core LMTs, or any learning-related increase in filopodia numbers at LMTs in CA3 (Fig. 4c). Furthermore, analysis of c-Fos-positive neurons upon recall 1 day after learning revealed a complete absence of ensemble activity rearrangements in CA3 upon fear conditioning, leading to comparable contents of c-Fos-positive neurons upon re-exposure to the training context or exposure to an unrelated neutral context, regardless of associative learning through aversive pairing (Fig. 4d). These results indicate that synaptic plasticity at LMTs in CA3 is required to re-organize pyramidal neuron ensemble activity in CA3 upon fear conditioning learning, to establish a precise memory of context in the hippocampus, and to induce learning-related feedforward inhibition growth.

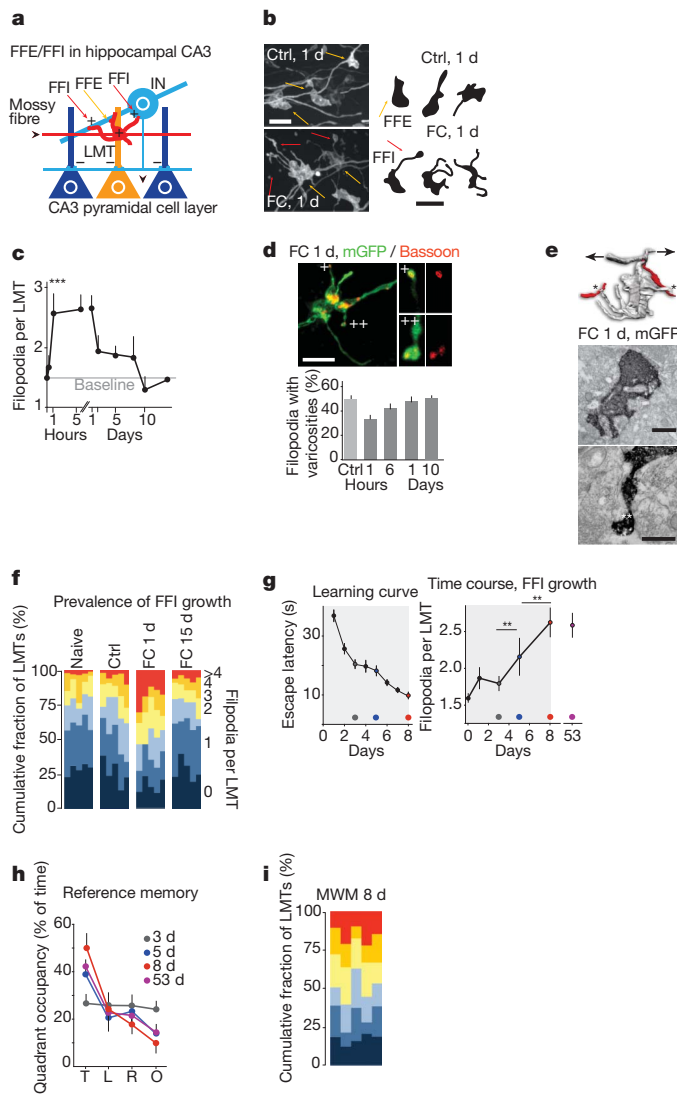


Figure 1 | Learning-related feedforward inhibition connectivity growth in the hippocampus. **a**, Schematic of hippocampal feedforward excitation (FFE) and feedforward inhibition (FFI) circuit in stratum lucidum of CA3. IN, inhibitory interneuron. **b–f**, Feedforward inhibition growth at hippocampal mossy fibre LMTs upon contextual fear conditioning. **b**, Micrographs and representative camera lucidas of mGFP-labelled mossy fibres and LMTs in hippocampal stratum lucidum (CA3b). Yellow arrows, core LMTs; red arrows, filopodia. Ctrl, control; FC, fear conditioning. **c**, Average filopodia/LMT values upon fear conditioning. $N = 5$ mice (100 LMTs each). $***P < 0.001$. **d**, Filopodial synapses upon fear conditioning. Overview panel shows maximal intensity projection of mGFP-positive LMT with four filopodia. Detail panels show single confocal planes of two of the filopodia (+ and ++); Bassoon channel masked using three-dimensional isosurface of GFP-positive LMT. Bar diagram shows fraction of LMT filopodia with varicosities as a function of time upon fear conditioning ($N = 3$, 100 LMTs). **e**, Filopodia upon fear conditioning learning contact spine-free dendrites. Immuno-electron microscopy of mGFP-positive LMT with four filopodia, 1 day after fear conditioning. Top, three-dimensional reconstruction of immuno-labelled LMT (red, spine-free dendrites contacted by two of the filopodia in the example (marked by one and two asterisks, respectively)). Centre, immuno-labelled LMT. Bottom, filopodium with contact is marked by two asterisks. **f**, Distributions of filopodia per LMT contents for individual mice. $N = 100$ LMTs. Relative contents of LMTs with 0, 1, 2, 3, 4, >4 filopodia as a fraction of the total LMT population. Vertical rows, individual mice. **g–i**, Feedforward inhibition growth at hippocampal mossy fibre LMTs upon Morris water maze (MWM) training. **g**, Learning curve and time course of feedforward inhibition growth. $N = 5$ mice (100 LMTs each). Grey area shows daily training period. The circles highlight the positions on the curves as compared to reference memory (right). **h**, Reference memory at 3, 5, 8 and 53 days. Percentage of time spent by the mice in target (T), left (L), right (C) and opposite (O) quadrants. $N = 5$ mice. **i**, Filopodia per LMT distributions after 8 days of training, as described in **b**. Scale bars, 5 μm (**b**, **d**, top and **g**), 1 μm (**d**, bottom centre) and 0.5 μm (**d**, bottom right).

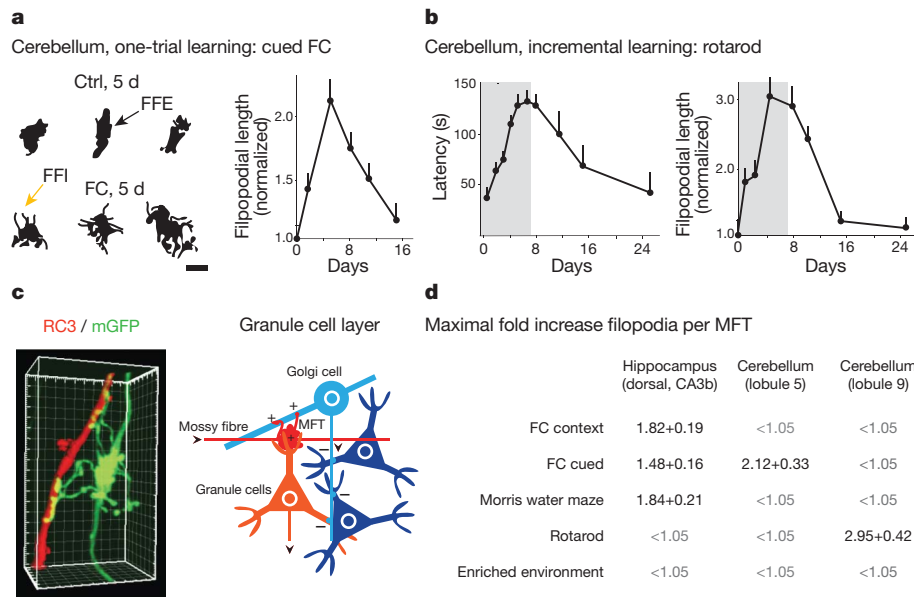


Figure 2 | Specificity of learning-related feedforward inhibition growth. **a, b,** Learning-related feedforward inhibition connectivity growth in cerebellar cortex. **a,** Feedforward inhibition growth at lobule 5 cerebellar cortex mossy fibre terminals upon cued fear conditioning. Labelling as in Fig. 1b. Scale bar, 10 μ m. **b,** Feedforward inhibition growth at lobule 9 cerebellar cortex mossy fibre terminals (MFTs) upon rotarod learning. Labelling as in Fig. 1c. **c,** In cerebellar cortex, mossy fibre terminal filopodia contact inhibitory Golgi cells. Left, three-dimensional rendering of contacts by mossy fibre terminal filopodia onto RC3-positive Golgi cell dendrite. Right, feedforward excitation/feedforward inhibition circuit in granule cell layer of cerebellar cortex. **d,** Specific relationship between learning and feedforward inhibition growth. Average fold increase values at peak response (fear conditioning hippocampus, 1 day; fear conditioning cerebellum, 2 days; Morris water maze, 8 days; rotarod, 5 days). $N = 5$, 100 LMTs or mossy fibre terminals each.

To test the notion that learning-related feedforward inhibition growth is necessary for memory precision, we then carried out learning experiments in *Add2* knockout mice¹⁰, which exhibit early LTP, but have a defect in synapse stabilization due to impaired linkage between the cell membrane cortex and the actin cytoskeleton²⁴. In naive *Add2*^{-/-} mice, average values of filopodia per LMT were closely comparable to those in wild-type mice. Unlike *Rab3a*^{-/-} mice, *Add2*^{-/-} mice did exhibit enhanced putative release sites per core LMT upon fear conditioning (Supplementary Material), but they completely failed to establish higher numbers of filopodia upon fear conditioning (Fig. 5a). In parallel, and like *Rab3a*^{-/-} mice, *Add2*^{-/-} mice learned to

associate fear with context, but the memory was imprecise and mice already generalized 1 day after fear conditioning (Fig. 5a). Comparable findings were obtained for Morris water maze and rotarod learning in *Add2*^{-/-} mice (Fig. 5b and Supplementary Material). Absence of learning-related feedforward inhibition connectivity growth in *Add2*^{-/-} mice is thus correlated with poor precision of the learned memory in the fear conditioning and Morris water maze paradigms, and with a near to complete failure to learn the rotarod task.

We then investigated c-Fos-positive CA3 pyramidal neuron ensembles in response to fear conditioning in the *Add2*^{-/-} mice. In stark contrast to *Rab3a*^{-/-} mice lacking mossy fibre LTP, and consistent with increased feedforward excitation connectivity, *Add2*^{-/-} mice exhibited c-Fos ensemble reorganization responses in CA3 that were qualitatively closely comparable to those in wild-type mice (Fig. 5c). Remarkably, however, net total numbers of c-Fos-positive neurons were more than 2.5 times higher for each experimental condition in *Add2*^{-/-} mice compared to wild-type mice (Fig. 5c). By contrast, numbers of c-Fos-positive pyramidal neurons in naive *Add2*^{-/-} mice were not higher than those in naive wild-type mice, indicating that the mutant mice did not just exhibit raised levels of c-Fos in CA3 neurons (Fig. 5c).

In wild-type mice, a reorganization of training context/neutral context ensembles upon fear conditioning was also detected in granule cells, but it was much less marked than in CA3 (Fig. 5d). Notably, however, and in stark contrast to CA3, distributions and numbers of c-Fos positive granule cells in *Add2*^{-/-} mice were not different from those in wild-type mice for all experimental conditions tested (Fig. 5d). Therefore, *Add2*^{-/-} mice re-organized their CA3 pyramidal neuron ensembles like wild-type mice, but failed to restrict the numbers of

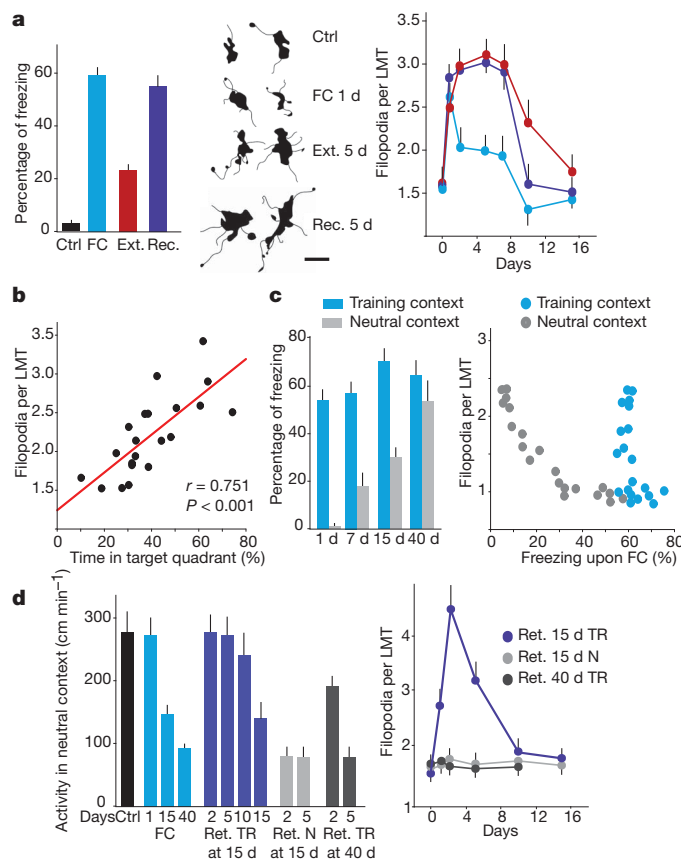


Figure 3 | Correlation between feedforward inhibition growth and quality of hippocampal learning and memory. **a,** Memory retrieval prolongs peak levels of feedforward inhibition growth upon cued fear conditioning (FC). Pale blue: fear conditioning, no recall (at 1 day); red: fear conditioning followed by extinction (Ext.) at 5 h and 24 h (at 5 days); violet: fear conditioning followed by recall (Ret.) at 5 h and 24 h (at 5 days). $N = 5$ mice (100 LMTs each). Scale bar, 5 μ m. **b,** Correlation between reference memory precision and average filopodial contents per LMT in Morris water maze task. Dots show individual mice analysed between day 1 and day 8 of the training procedure (100 LMTs each). **c,** Time-dependent generalization upon contextual fear conditioning learning. Right, dots represent average values for individual mice at different times after fear conditioning learning (100 LMTs each). **d,** Re-growth of filopodia and re-contextualization upon retrieval of training context memory (Ret. TR) versus retrieval of neutral context (Ret. N). Left, exploratory activity in neutral context as a function of days after last manipulation. Error bars show mean \pm s.e.m.

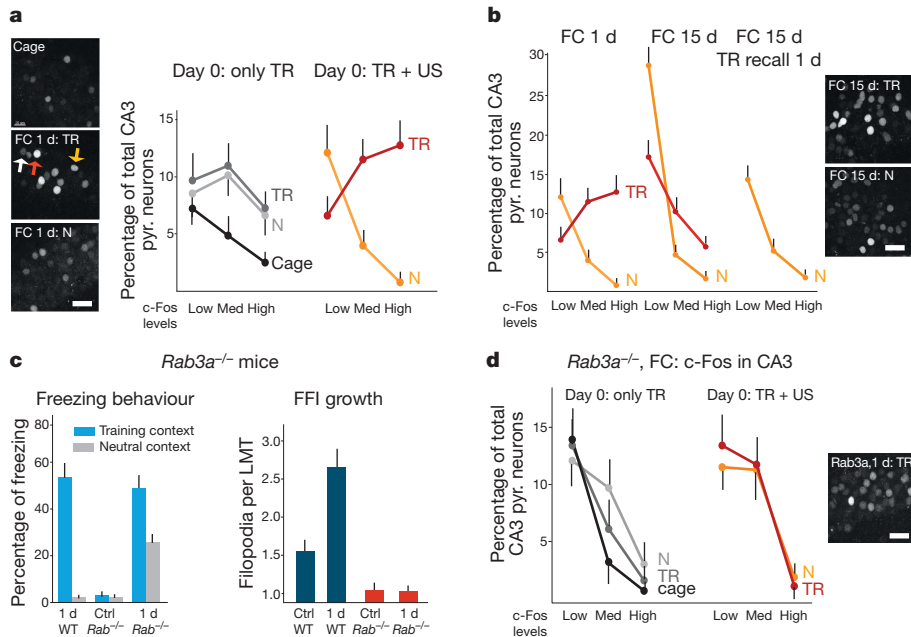


Figure 4 | Relationship between induction of c-Fos in CA3 pyramidal neurons and behavioural memory precision upon contextual fear conditioning. **a**, c-Fos immunoreactivity in CA3 pyramidal neurons upon exposure to training context (TR) or neutral context (N), with or without aversive association. Panels show representative examples of c-Fos immunoreactivity in CA3b. c-Fos neurons classified as weak (white arrow), medium (yellow arrow), strong (red arrow). *N* = 3, 500 pyramidal (pyr.) neurons each. US, unconditioned, aversive stimulus. **b**, c-Fos immunoreactivity in CA3 pyramidal neurons 15 days after fear conditioning: effect of recall with training context. Details as in **a**. *N* = 1,000–1,500 pyramidal neurons, from 3 mice each. **c**, Generalization and absence of learning-induced feedforward inhibition growth in *Rab3a*^{-/-} mice. *N* = 5 mice (100 LMTs each). **d**, c-Fos immunoreactivity in CA3 pyramidal neurons of *Rab3a*^{-/-} mice upon fear conditioning. Details as in **a**. *N* = 1,000–1,500 pyramidal neurons, from 3 mice each. Scale bars, 20 μ m. Error bars show mean \pm s.e.m.

activated pyramidal neurons in CA3 upon stimuli, which is consistent with a complete absence of feedforward inhibition connectivity growth at LMTs. Furthermore, c-Fos activation patterns in CA3 correlated with memory precision, whereas those in dentate gyrus did not, suggesting that the absence of Add2 in mossy fibres and their LMTs may account for the impaired memory precision in *Add2*^{-/-} mice.

To establish a causal link between learning-related feedforward inhibition growth at LMTs and memory precision, we determined whether re-expression of Add2 specifically in granule cells and their mossy fibres was sufficient to rescue filopodial growth and memory precision upon fear conditioning. To achieve specific re-expression in the adult, we expressed Add2 selectively in the dentate gyrus¹⁵ of

Add2^{-/-} mice using a lentiviral construct. One month after viral transduction, 15–22% of granule cells throughout the entire hippocampus exhibited virus-driven gene expression, whereas expression outside the dentate gyrus was extremely rare (Fig. 5e). The re-introduction of Add2 in mossy fibres was sufficient to rescue filopodial growth at LMTs of transduced granule cells in response to fear conditioning (Fig. 5f). Most notably, and in parallel to restored feedforward inhibition growth, re-expression of Add2 in granule cells rescued behavioural contextualization upon fear conditioning (Fig. 5g).

Our results establish a causal relationship between learning-associated structural alterations in identified circuit connectivity and a specific behavioural output. We provide evidence that increased feedforward

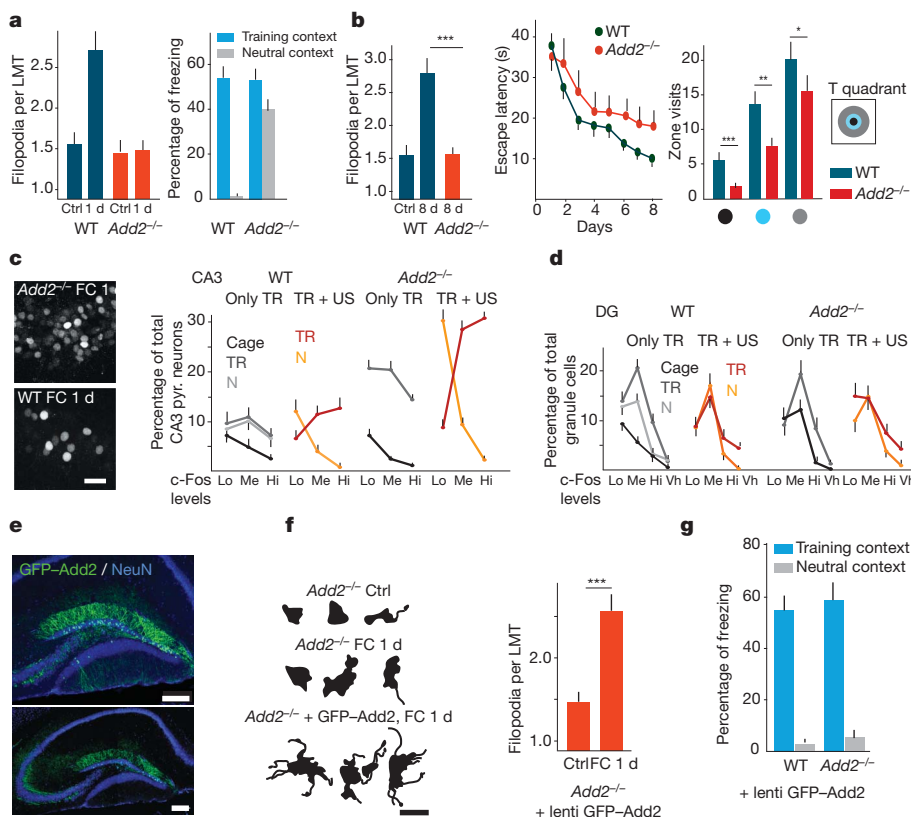


Figure 5 | Critical role of mossy fibre Add2 for feedforward inhibition growth at LMTs and hippocampal memory precision. **a**, Absence of feedforward inhibition growth upon contextual fear conditioning, and generalization in *Add2*^{-/-} mice. Conditions as in Fig. 4c. **b**, Absence of feedforward inhibition growth upon Morris water maze learning, and imprecise spatial memory in *Add2*^{-/-} mice. Conditions as in Fig. 4c. **c**, **d**, c-Fos immunoreactivity in CA3 pyramidal neurons (**c**) and in dentate gyrus (DG) granule cells (**d**) of wild-type (WT) and *Add2*^{-/-} mice upon exposure to training context (TR) or neutral context (N), with or without aversive conditioning. Hi, high; Lo, low; Me, medium; Vh, very high. Conditions as in Fig. 4a. **e–g**, Rescue of feedforward inhibition growth and contextualization upon re-expression of Add2 in granule cells of adult *Add2*^{-/-} mice. **e**, Examples of transduced hippocampus (dorsal third of hippocampus). **f**, Rescue of feedforward inhibition growth; luciferase: transgene expression visualized by the GFP-Add2 construct in the absence of mGFP reporter. **g**, Behavioural rescue of contextualization. Conditions as in **a**. Scale bars: 5 μ m (**f**), 20 μ m (**c**) and 200 μ m (**e**). Error bars show mean \pm s.e.m.

inhibition connectivity upon learning by mossy fibre LMTs in CA3 is critically important for the behavioural precision of learning-related hippocampal spatial memories. We further show that, upon learning, the increased feedforward inhibition connectivity is brought about through structural plasticity at a substantial fraction of LMTs in CA3, leading to about a doubling in the numbers of excitatory synapses onto parvalbumin-positive inhibitory interneurons (see also Supplementary Material).

Our results introduce a distinction between spatial learning, which is present in *Add2*^{-/-} mice, and the behavioural precision of the learning, which is compromised in these mutant mice. The distinction is consistent with the notion that the hippocampus is critically important for the precision of contextual memories²⁵. Within the hippocampal circuit, the dentate gyrus establishes fine-grained representations of experience, which it transmits to CA3 (ref. 13). Upon learning-induced potentiation, this high-resolution information may augment the detection of similarities among unrelated events through the associational network in CA3. Accordingly, filtering of the mossy fibre output through feedforward inhibition connectivity upon learning^{26–28} may support memory precision by restricting the extraction of relational representations in CA3 (ref. 29). The increase in feedforward inhibition connectivity through structural plasticity discovered in this study may thus have important roles in ensuring the precision of behaviourally relevant memories upon learning, under normal and pathological conditions.

METHODS SUMMARY

Rab3a^{-/-} and *Add2*^{-/-} mice^{9,10} were from Jackson Laboratories; the reporter line *Thy1-mGFP(Lsl1)* was as described before⁵. The membrane-targeted green fluorescent protein (mGFP) lentivirus to trace mossy fibre projections was as described previously¹⁵; the GFP-Add2 construct was cloned into a lentivirus vector, and dentate gyrus infections were as described previously¹⁵.

For anatomical analysis, mice were perfused with ice-chilled 4% paraformaldehyde in 0.1 M PBS, and brains were post-fixed. Hippocampi were mounted in 3% agarose blocks, and 100-µm transversal sections of hippocampi were cut using a McIlwain tissue chopper. Sections analysed were within 15% and 30% along the anterior–posterior axis. All LMTs that could be resolved in three dimensions within any given optical field (×100) were analysed for filopodial contents. Filopodia were defined as processes emanating from LMTs of at least 2 µm length; varicosities were defined as end-swellings of at least 1 µm in diameter.

The immuno-electron microscopy analysis was performed according to a published procedure³⁰.

For c-Fos analysis, mice were perfused for 90 min after the last memory recall. Quantitative analysis of Bassoon puncta and c-Fos-positive nuclei was performed using a computerized image analysis system (Imaris 7, Bitplane). Nuclei were detected automatically as spheres of 8 µm, and the software yielded distributions of c-Fos-positive nuclei. Intensity thresholds for CA3 were defined as follows: low (>280, <450), medium (>450, <700), high (>700; the highest values were about 1,400).

Statistical analyses were performed using Student's *t*-tests and one-way ANOVA; post hoc comparisons were at the *P* < 0.05 level of significance. Results are presented as mean ± s.e.m.

All behavioural experiments were carried out with male mice that were 55–65 days old at the onset of the experiment, and were according to standard procedures. All subsequent morphological and immunohistochemical analyses of behaviourally treated mice were carried out blind to behavioural conditions.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 September 2010; accepted 17 February 2011.

Published online 1 May 2011.

- Holtmaat, A. & Svoboda, K. Experience-dependent structural plasticity in the mammalian brain. *Nature Rev. Neurosci.* **10**, 647–658 (2009).
- Hübener, M. & Bonhoeffer, T. Searching for engrams. *Neuron* **67**, 363–371 (2010).
- Lamprecht, R. & LeDoux, J. Structural plasticity and memory. *Nature Rev. Neurosci.* **5**, 45–54 (2004).
- Wilbrecht, L., Holtmaat, A., Wright, N., Fox, K. & Svoboda, K. Structural plasticity underlies experience-dependent functional plasticity of cortical circuits. *J. Neurosci.* **30**, 4927–4932 (2010).
- De Paola, V., Arber, S. & Caroni, P. AMPA receptors regulate dynamic equilibrium of presynaptic terminals in mature hippocampal networks. *Nature Neurosci.* **6**, 491–500 (2003).

- Hofer, S. B., Mrcic-Flogel, T. D., Bonhoeffer, T. & Hübener, M. Experience leaves a lasting structural trace in cortical circuits. *Nature* **457**, 313–317 (2009).
- Xu, T. *et al.* Rapid formation and selective stabilization of enduring motor memories. *Nature* **462**, 915–919 (2009).
- Yang, G., Pan, F. & Gan, W. B. Stably maintained dendritic spines are associated with lifelong memories. *Nature* **462**, 920–924 (2009).
- Castillo, P. E. *et al.* Rab3A is essential for mossy fibre long-term potentiation in the hippocampus. *Nature* **388**, 590–593 (1997).
- Rabenstein, R. L. *et al.* Impaired synaptic plasticity and learning in mice lacking β-adducin, an actin-regulating protein. *J. Neurosci.* **25**, 2138–2145 (2005).
- Wang, S.-H. & Morris, R. G. Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation. *Annu. Rev. Psychol.* **61**, 49–79 (2010).
- Nakashiba, T., Young, J. Z., McHugh, T. J., Buhl, D. L. & Tonegawa, S. Transgenic inhibition of synaptic transmission reveals role of CA3 output in hippocampal learning. *Science* **319**, 1260–1264 (2008).
- Leutgeb, J. K., Leutgeb, S., Moser, M. B. & Moser, E. I. Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science* **315**, 961–966 (2007).
- Gogolla, N., Galimberti, I., Deguchi, Y., Caroni, P. Wnt signaling mediates experience-related regulation of synapse numbers and mossy fiber connectivities in the hippocampus. *Neuron* **62**, 510–525 (2009).
- Galimberti, I., Bednarek, E., Donato, F. & Caroni, P. EphA4 signaling in juveniles establishes topographic specificity of structural plasticity in the hippocampus. *Neuron* **65**, 627–642 (2010).
- Acsády, L., Kamondi, A., Sik, A., Freund, T. & Buszaki, G. GABAergic cells are the major postsynaptic target of mossy fibers in the rat hippocampus. *J. Neurosci.* **18**, 3386–3403 (1998).
- Lawrence, J. J. & McBain, C. J. Interneuron diversity series: containing the detonation—feedforward inhibition in the CA3 hippocampus. *Trends Neurosci.* **26**, 631–640 (2003).
- Mori, M., Abegg, M. H., Gaehwiler, B. H. & Gerber, U. A frequency-dependent switch from inhibition to excitation in a hippocampal unitary circuit. *Nature* **431**, 453–456 (2004).
- D'Angelo, E. & De Zeeuw, C. I. Timing and plasticity in the cerebellum: focus on the granular layer. *Trends Neurosci.* **32**, 30–40 (2009).
- Sacchetti, B., Scelfo, B., Tempia, F. & Strata, P. Long-term synaptic changes induced in the cerebellar cortex by fear conditioning. *Neuron* **42**, 973–982 (2004).
- Wiltgen, B. J. & Silva, A. J. Memory for context becomes less specific with time. *Learn. Mem.* **14**, 313–317 (2007).
- Biedenkapp, J. C. & Rudy, J. W. Context pre-exposure prevents forgetting of a contextual fear memory: implication for regional changes in brain activation patterns associated with remote and recent memory tests. *Learn. Mem.* **14**, 200–203 (2007).
- Kubik, S., Miyashita, T. & Guzowski, J. F. Using immediate-early genes to map hippocampal subregional functions. *Learn. Mem.* **14**, 758–770 (2007).
- Bednarek, E. & Caroni, P. β-Adducin is required for stable assembly of new synapses and improved memory upon environmental enrichment. *Neuron*. (in the press).
- Wiltgen, B. J. *et al.* The hippocampus plays a selective role in the retrieval of detailed contextual memories. *Curr. Biol.* **20**, 1336–1344 (2010).
- Lamsa, K., Heeroma, J. H. & Kullmann, D. M. Hebbian LTP in feed-forward inhibitory interneurons and the temporal fidelity of input discrimination. *Nature Neurosci.* **8**, 916–924 (2005).
- Wulff, P. *et al.* Synaptic inhibition of Purkinje cells mediates consolidation of vestibulo-cerebellar motor learning. *Nature Neurosci.* **12**, 1042–1049 (2009).
- Pouille, F., Marin-Burgin, A., Adesnik, H., Atallah, B. V. & Scanziani, M. Input normalization by global feedforward inhibition expands cortical dynamic range. *Nature Neurosci.* **12**, 1577–1585 (2009).
- McNaughton, B. L. & Morris, R. G. M. Hippocampal synaptic enhancement and information storage within a distributed memory system. *Trends Neurosci.* **10**, 408–415 (1987).
- Knott, G. W., Holtmaat, A., Trachtenberg, J. T., Svoboda, K. & Welker, E. A protocol for preparing GFP-labeled neurons previously imaged *in vivo* and in slice preparations for light and electron microscopic analysis. *Nature Protocols* **4**, 1145–1156 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Arber and B. Roska for valuable comments on the manuscript. We are grateful to J. Pielage for sharing with us his findings on the function of Add2 in synapse stability, and to G. Courtine for advice on the c-Fos labelling protocol. The Friedrich Miescher Institut is part of the Novartis Research Foundation.

Author Contributions S.R. devised, carried out and analysed all experiments except for those of Fig. 2a–c, part of Fig. 2d, Fig. 5a, e–g and Supplementary Fig. 4; C.V. carried out the experiments of Fig. 2a–c, part of Fig. 2d and Supplementary Fig. 4; E.B. devised and carried out the behavioural and rescue experiments on *Add2*^{-/-} mice; C.G. carried out the immuno-electron microscopy experiments; B.S. provided advice in planning and interpreting the fear conditioning experiments; P.S. provided advice on the cerebellar experiments; P.C. helped devise the experiments and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.C. (caroni@fmi.ch).

METHODS

Reagents and immunocytochemistry. Antibodies were from the following sources, and were used as follows: parvalbumin, Swant, 1:5,000; VGluT1, SySy, 1:1,000; GAD65/67, Millipore, 1:1,000; c-Fos, Santa Cruz, 1:10,000; NeuN, Chemicon, 1:200; Bassoon, Millipore, 1:200; Alexa-labelled secondary antibodies, Molecular Probes, 1:500.

For immunocytochemistry, tissues were permeabilized with 0.2% Triton X-100 in PBS with 10% bovine serum albumin (BSA). Antibody incubations were overnight at 4 °C.

Fluorescence was imaged on either an upright spinning disk microscope consisting of a Yokogawa CSU22 confocal scanning head mounted on a Zeiss Axiomager M1 using a $\times 100$ alphaPlan-Apochromat 1.45 (Zeiss) oil-immersion objective, or on an LSM510 confocal microscope (Zeiss) using a $\times 63$ (1.4) oil-immersion objective.

At least four sections were analysed per mouse, and the data are based on 300–500 μm regions along the anterior–posterior axis.

c-Fos analysis. For c-Fos analysis, all samples belonging to the same experimental set were processed in parallel. Occasional sections in which NeuN signals were lower than average, or where c-Fos signal intensities varied within different regions of the section were discarded as technically poor. All images were acquired with the same settings, which were defined in order to avoid saturation of the highest c-Fos signals in CA3 and dentate gyrus, and to still detect background levels outside cell clusters. Cells were binned according to labelling intensities using an automatic procedure, and the same threshold settings were used for all experiments. For dentate gyrus granule cells, the thresholds were as follows: low (>280 , <450), medium (>450 , <700), high (>700 , $<1,000$), very high ($>1,000$; the highest values were about 2,200). c-Fos immunoreactive neurons were counted using a minimum of four sections per animal, and normalized to the total number of NeuN-positive nuclei within the neuronal layers in CA3 or dentate gyrus. In a first series of experiments, batches of naive and fear conditioning control mice (training context without unconditioned aversive stimulus) were tested for inter-animal variability, which was found to be very low.

Behavioural experiments. The behavioural experiments were in accordance with institutional guidelines, and were approved by the Veterinary Department of the Canton of Basel-Stadt. Mice were kept in temperature-controlled rooms on a constant 12 h light/dark cycle, and experiments were conducted at the approximate same time during the light cycle. Before the behavioural experiments, mice were kept in a holding room in single cages for 3–4 days. At the onset of each behavioural experiment mice were 50–60 days old.

For the Morris water maze test, the 140 cm pool was surrounded by black curtains, and by four different objects. A circular escape platform (10 cm diameter) was submerged 0.5 cm below the water surface, and was kept in a fixed position. Mice were trained to find the platform for 4 trials a day, during up to 8 days. During training, mice were released from pseudo-randomly assigned start locations; they were allowed to swim for up to 60 s, when they were manually guided to the platform in the case of failures. Inter-trial intervals were 5 min. Single probe trials to test reference memory were conducted 1 day after the last training session. Mice were released at a random start position, and were allowed to swim during 60 s in the absence of the platform.

The training context (TR) was rectangular, and was cleaned with 1% acetic acid before and after each trial; the neutral context (N) had a cylindrical shape and was cleaned with 70% ethanol. Freezing was defined as the absence of somatic motility, except for respiratory movements. Exploratory activity was measured as body distance travelled over time. Once placed in the conditioning chamber, the mice were allowed to freely explore for 2.5 min, and they received 5 presentation of conditioned stimulus and unconditioned stimulus (1 s foot shock, 0.8 mA; where indicated, 10 kHz tone for 10 s, 70 dB sound pressure level, inter-trial interval 30 s). The last 1 s of each tone was paired with the unconditioned stimulus. Contextual fear conditioning involved the same protocol, but without the tone component. To test for contextual fear memory, mice were returned to training (or neutral) context during a test period of 2.5 min. To test for cued fear conditioning, mice explored for 2 min, followed by 5 tone presentations. The test was performed either in the conditioning context (context- and tone-dependent freezing), or in a novel context (tone-dependent freezing).

To test for context discrimination after fear conditioning, a within-subjects design was used. On the test day, freezing was assessed in training context during 2.5 min, and 5 h later in neutral context. Where indicated, mice were tested for generalization in neutral context, followed 5 h and 24 h later by two brief recall sessions (in training or neutral context). Subsequently, discrimination was tested in a second novel context (novel room shape; 0.25% benzaldehyde/ethanol).

Data from training sessions and probe trials were collected and analysed using Viewer2 Software (Bioobserve). Cued and contextual fear conditioning were carried out in the Mouse Test Cage (Coulbourn Instruments). Freezing behaviour was scored using Ethovision software (Noldus). Mice were excluded from the data set if they failed at the behavioural analysis; this was the case when mice failed to extinguish fear responses to training context (two mice), exhibited weak freezing to training context in the recall experiments at day 15 (three mice), exhibited signs of behavioural extinction upon recall of training context at day 15 (seven mice), or failed to learn the Morris water maze (one mouse).

Transmission electron microscopy. This procedure is described in detail elsewhere³⁰. Briefly, mice were transcardially perfused with 2% paraformaldehyde and 0.2% glutaraldehyde in PBS 0.1 M pH 7.4. Right and left hippocampi were dissected, and 60 μm vibratome (Leica) sections were obtained, rinsed, cryoprotected and freeze-thawed in liquid nitrogen. Sections were incubated in first antibody (GFP, chemicon 1:1,000) overnight, followed by biotinylated secondary antibody (Invitrogen 1:500). After incubation in the avidin-biotin peroxidase complex (ABC elite, Vector Laboratories), labelling was performed with DAB and hydrogen peroxide. After the revelation of the labelling, sections were stained in osmium tetroxide and dehydrated. After impregnation with Durcupan resin (FLUKA) sections were flat-embedded between two silicon-coated glass slides and cured in a 60 °C oven for 48 h.

Transmission light microscopy was performed in stratum lucidum to search for large mossy fibre terminals with more than three filopodia. Appropriate blocks were then trimmed, and 60 nm serial sections were cut and collected on formvar coated slot-grids. Images of labelled terminal were acquired with a side-mounted digital camera (Veleta, Olympus) on a Philips CM10 transmission electron microscopy at 80 kV, and a pixel size of 2.63 nm. To reconstruct the structure in three dimensions, images were aligned (Autoaligner, Bitplane), and contours were drawn manually using Imaris 7.1.2 (Bitplane). Surface rendering was achieved using Geometry converter (J. Wolf) and Blender.

Improved molecular replacement by density- and energy-guided protein structure optimization

Frank DiMaio¹, Thomas C. Terwilliger², Randy J. Read³, Alexander Wlodawer⁴, Gustav Oberdorfer⁵, Ulrike Wagner⁵, Eugene Valkov⁶, Assaf Alon⁷, Deborah Fass⁷, Herbert L. Axelrod⁸, Debanu Das⁸, Sergey M. Vorobiev⁹, Hideo Iwai¹⁰, P. Raj Pokkuluri¹¹ & David Baker¹

Molecular replacement^{1–4} procedures, which search for placements of a starting model within the crystallographic unit cell that best account for the measured diffraction amplitudes, followed by automatic chain tracing methods^{5–8}, have allowed the rapid solution of large numbers of protein crystal structures. Despite extensive work^{9–14}, molecular replacement or the subsequent rebuilding usually fail with more divergent starting models based on remote homologues with less than 30% sequence identity. Here we show that this limitation can be substantially reduced by combining algorithms for protein structure modelling with those developed for crystallographic structure determination. An approach integrating Rosetta structure modelling with Autobuild chain tracing yielded high-resolution structures for 8 of 13 X-ray diffraction data sets that could not be solved in the laboratories of expert crystallographers and that remained unsolved after application of an extensive array of alternative approaches. We estimate that the new method should allow rapid structure determination without experimental phase information for over half the cases where current methods fail, given diffraction data sets of better than 3.2 Å resolution, four or fewer copies in the asymmetric unit, and the availability of structures of homologous proteins with >20% sequence identity.

The limiting steps in molecular replacement are finding the correct location of the starting model in the unit cell and the interpretation of electron density maps produced using the imperfect phase information from candidate model placements. The left column of Fig. 1 illustrates the problem of initial model-building starting with distant comparative models (20–30% sequence identity) that have been correctly placed in the crystallographic unit cell. Automatic chain tracing methods fail on such maps because they often follow the incorrect comparative model (red) more closely than the actual structure (yellow); breaks in the density make it difficult to recover the correct backbone trace. Nevertheless, the maps contain considerable information about the native structure; for example, portions of the starting model that are not within density are generally incorrect.

Structure prediction methods such as Rosetta search for the lowest energy conformation of the polypeptide chain using physically realistic force fields. Based on previous work showing that accurate structures could be obtained from very sparse NMR data sets¹⁵ by using the data to guide structure prediction searches, we reasoned that structure prediction methods guided by even very noisy density maps might be able to improve a poor molecular replacement model before applying crystallographic model-building techniques. We developed an approach in which electron density maps generated from molecular replacement solutions for each of a series of starting models are used to guide energy optimization by structure rebuilding, combinatorial side chain packing, and torsion space minimization¹⁶. New maps are generated using phase information from the energy-optimized models

most consistent with the diffraction data, subjected to automatic chain tracing, and success is monitored through the free *R* factor¹⁷.

To investigate the performance of the new method, we obtained 18 crystallographic data sets that had resisted previous attempts at structure determination. We first tested whether a comprehensive set of state-of-the-art molecular replacement approaches using a range of full-length and trimmed templates and homology models could solve any of these structures (Supplementary Information). We were able to solve five of the structures with both the new method and the existing methods (Table 1), leaving 13 challenging data sets highly resistant (Supplementary Information section 1) to structure determination (Table 1). For each of these, we identified homologous proteins of known structure¹⁸ and constructed sequence alignments and starting models⁹ from the five closest homologues. Starting models were used to search for up to five

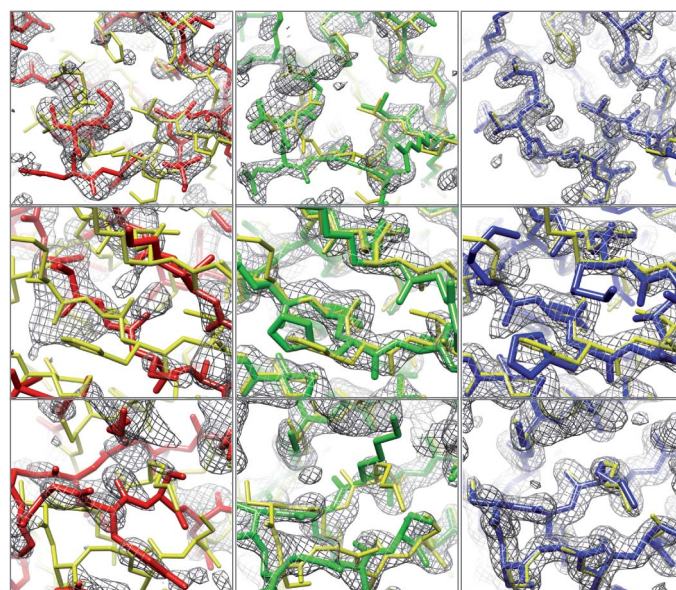


Figure 1 | Examples of improvement in electron density and model quality. Each row corresponds to one of the entries in Table 1. First row: 6 (2.0 Å resolution); second row: 7 (2.1 Å resolution); third row: 12 (1.7 Å resolution). Left column: correct initial molecular replacement solution (not necessarily identifiable at this stage) using starting model and corresponding density. Middle column: energy-optimized model and corresponding density. Right column: model and density following automatic building using the energy-optimized model as the source of phase information. The final deposited structure is shown in yellow in each panel; the initial model, energy-optimized model, and model after chain rebuilding are in red, green and blue, respectively. The sigma-A-weighted $2mF_o - DF_c$ density contoured at 1.5σ is shown in grey.

¹University of Washington, Department of Biochemistry and HHMI, Seattle, Washington 98195, USA. ²Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. ³University of Cambridge, Department of Haematology, Cambridge Institute for Medical Research, Cambridge CB2 0XY, UK. ⁴Macromolecular Crystallography Laboratory, National Cancer Institute at Frederick, Frederick, Maryland 21702, USA. ⁵Institute of Molecular Biosciences, University of Graz, Humboldtstrasse 50/3, 8010-Graz, Austria. ⁶University of Cambridge, Department of Biochemistry, Cambridge CB2 1GA, UK. ⁷Weizmann Institute of Science, Department of Structural Biology, Rehovot 76100, Israel. ⁸Joint Center for Structural Genomics and SSRL, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA. ⁹Northeast Structural Genomics Consortium, Columbia University, New York, New York 10027, USA. ¹⁰University of Helsinki, Institute of Biotechnology, FI-00014 Helsinki, Finland. ¹¹Argonne National Laboratory, Biosciences Division, Argonne, Illinois 60439, USA.

Table 1 | Determination of previously unsolved structures using the new approach

R _{free} after Phaser MR and model-building protocol											
ID number	Source*	Resolution (Å)	Seqid (%)	Autobuild	Arp/Warp	Simulated annealing (SA) + Autobuild	Torsion-space SA + Autobuild	Extreme SA + Autobuild	DEN + Autobuild	Rosetta + Autobuild	R _{free} (current best)
Solved by multiple methods											
1	JCSG	2.1	22	0.31	0.50	0.30	0.30	0.30 †	0.35	0.31	0.22
2	NSGC	2.2	19	0.29	0.57	0.29	0.29	0.29 †	0.30	0.29	0.22
3	UG	2.5	27	0.34	0.59	0.29	0.29	0.29 †	0.35	0.27	0.19
4	JCSG	2.7	21	0.31	0.59	0.30	0.30	0.30 †	0.31	0.30	0.24
5	ANL	1.9	31	0.51	0.59	0.54	0.54	0.24	0.39	0.31	0.24
Only solved by Rosetta											
Rosetta modelling with density required for successful model-building											
6	NCI	2.0	30	0.56	0.59	0.60	0.55	0.55	0.50	0.34	0.20
7	WI	2.1	22/15	0.56	0.60	0.54	0.54	0.54	0.56	0.28	0.26
8	JCSG	2.8	29	0.52	0.55	0.50	0.50	0.51	0.45	0.36	0.36‡
9	UC	3.0	22	0.54	0.56	0.50	0.50	0.47	0.46	0.32	0.25§
10	JCSG	3.2	20	0.54	0.57	0.51	0.51	0.53	0.46	0.39	0.33‡
11	UG	2.5	18	0.52	0.57	0.54	0.52	0.54	0.55	0.27	0.22
	MEAN			0.54	0.57	0.53	0.52	0.52	0.50	0.33	
Rosetta homology modelling required for successful molecular replacement											
12	BI, HY	1.7	— (100)	—	—	—	—	—	—	0.29	0.22
13	JCSG	2.9	29	—	—	—	—	—	—	0.39	0.23

The Seqid column gives the sequence identity to the closest homologue identified by HHpred¹⁸, and is shown in parentheses if this is an NMR structure. The next seven columns give the R_{free} of the model produced by different combinations of refinement and autobuilding approaches. The final column gives the R_{free} after further refinement by the crystallographer who provided the data. For structures solved by multiple methods, the new method as well as one or more alternative approaches was sufficient ($R_{\text{free}} < 0.4$). In the first subset of structures that could only be solved by the new method (only solved by Rosetta), molecular replacement succeeds (in some cases ambiguously) using the template alone but model-building fails; in the second subset, refinement in Rosetta is required for molecular replacement to succeed. Targets that could not be solved by our approach are listed in Supplementary Table 1.

*JCSG, Joint Center for Structural Genomics; NSGC, Northeast Center for Structural Genomics; UG, University of Graz; ANL, Argonne National Lab; NCI, National Cancer Institute; WI, Weizmann Institute of Science; UC, University of Cambridge; BI, HY, Institute Of Biotechnology, University of Helsinki.

† Because a single SA trajectory was sufficient to solve these cases, Extreme SA was not run. Values from the single SA run are shown for completeness.

‡ Solutions for both are essentially correct based on the selenium positions in the anomalous difference Fourier maps calculated from the experimental data. However, structures are difficult to complete to deposition due to some MR solution model bias, poor or disordered density in numerous regions and low resolution.

§ Refinement ongoing.

|| This structure was solved and all tests on this template were carried out using the intact template as a starting point. With this template both the molecular replacement step and subsequent rebuilding required Rosetta modelling for success. After determining the structure and completing the tests we found that it was also possible to solve the structure by molecular replacement if the template were split into two rigid subunits and the two domains were correctly chosen.

candidate molecular replacement solutions based on the likelihood of the experimental diffraction data². Electron density maps were computed for each of these solutions, and used to guide energy minimization by first remodelling the unaligned regions and regions which poorly fit the density and then optimizing all backbone and side chain torsion angles. The likelihood of the experimental diffraction data was computed for each optimized model²; if top ranked models were similar (see Methods), a map generated from the highest likelihood model was subjected to automatic chain rebuilding, density modification and refinement⁵. If this succeeded in building the majority of the protein and produced a model with free R factor¹⁷ significantly better than random ($R_{\text{free}} < 0.4$), the structure was considered solved; rebuilt models were further analysed by the crystallographers who supplied the original data. Using this approach, we were able to solve eight of the thirteen challenging cases (Table 1). In some of these eight cases, recognition of the correct placement of the model in the unit cell was only possible after Rosetta refinement (Supplementary Fig. 2); in others the correct placement was clear but the density was too poor for chain rebuilding. In two of the cases (12 and 13), even finding the correct molecular replacement solution first required energy-based refinement¹².

The improvement in electron density produced by density guided energy optimization and autobuilding are illustrated in Fig. 1. The starting molecular replacement models are often quite inaccurate, and the density generated from these models has breaks within the backbone of the actual structure (left panels). After model rebuilding and energy guided structure optimization, backbone breaks are largely closed and both side chains and backbone are more correctly modelled (middle panels). Automatic chain rebuilding into the improved map followed by density modification and reciprocal-space refinement further improve the model and the density (right panels). For all eight cases, the correlation between the final refined density and density from the original molecular replacement solutions is low, increases significantly after energy- and density-based structure optimization, and still further after automatic chain rebuilding (Supplementary Table 2).

For each of the eight challenging cases solved with the new method we also applied a battery of existing methods (Table 1 and

Supplementary Information section 1) including simulated annealing in Cartesian and torsion space in PHENIX and CNS¹⁴, deformable elastic network (DEN) refinement¹³ in CNS, and PHENIX Autobuild⁶ and ARP-WARP⁵ for model-building. As noted above, in two cases Rosetta structure modelling was required for the correct placement of starting models in the unit cell, so the alternative methods could not even be applied. In the remaining six cases, final R_{free} values were lower using the new approach than with any of the existing methods (Table 1, Fig. 2a). Whereas conventional simulated annealing in both Cartesian and torsion space had little effect, the recently developed DEN¹⁹ refinement protocol did improve three of the structures slightly, yielding free R values of 0.45–0.46 for these targets. Combination of DEN refinement with the method described here could lead to still more powerful approaches.

To benchmark the sequence and structural divergence where the different methods break down, we studied two different protein families for which a total of 59 different template structures covering a broad range of sequence and structural similarity were available (Supplementary Tables 3–5). Each template was correctly placed in the unit cell, and then improved with either Rosetta energy- and density-based optimization, Cartesian- and torsion-space simulated annealing, or DEN refinement. For each resulting model, the correlation with the density of the deposited structure was evaluated. Automatic chain rebuilding beginning with the superimposed starting models was successful for 18 of the 59 cases, consistent with the observation that molecular replacement often fails with templates sharing less than 30% sequence identity with the target sequence. Torsion-space simulated annealing in CNS before autobuilding allowed solution of two additional structures, DEN refinement, three additional structures, and Rosetta energy-based structure optimization, fourteen additional structures (Supplementary Fig. 2 and Supplementary Tables 3–5). We found the radius of convergence of the new method can be further extended by guiding energy based structure optimization by the Patterson correlation²⁰ rather than electron density (see Supplementary Information). This allowed structure improvement and identification of the correct molecular replacement solution in two additional cases (Supplementary Fig. 2, compare green

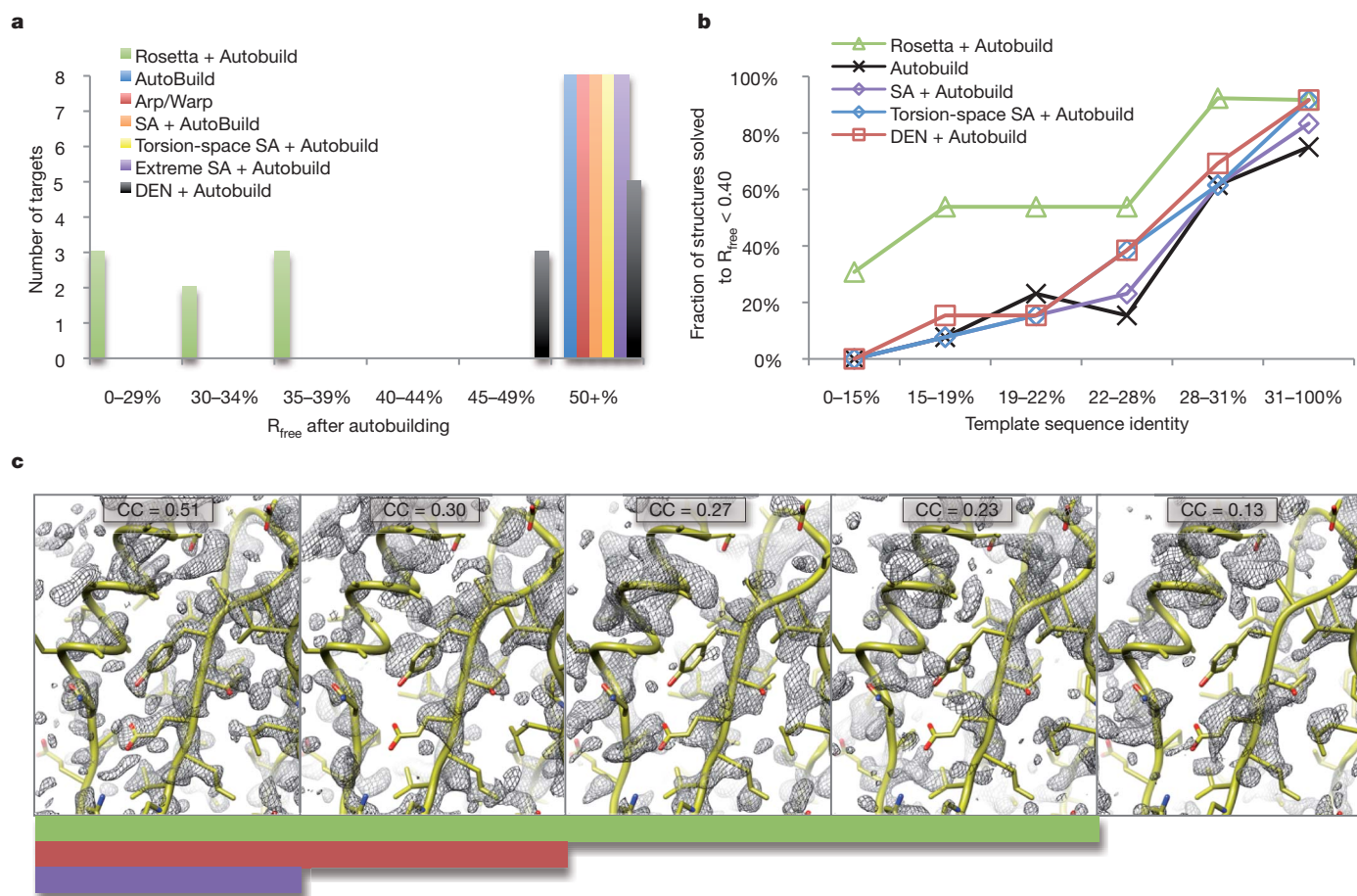


Figure 2 | Method comparison. **a**, Histogram of R_{free} values after autobuilding for the eight difficult blind cases solved using the new approach (Table 1). For most existing approaches, none of the cases yielded R_{free} values under 50%; DEN was able to reduce R_{free} to 45–49% for three of the structures. For all eight cases, Rosetta energy and density guided structure optimization led to R_{free} values under 40%. **b**, Dependence of success on sequence identity. The fraction of cases solved (R_{free} after autobuilding $< 40\%$) is shown as a function of template sequence identity over the 18 blind cases and 59 benchmark cases. The new method is a clear improvement below 28% sequence identity.

to orange bar); for one of these the improvements were sufficient for autobuilding to effectively solve the structure.

Over the combined set of 18 blind cases and the 59 benchmark cases, Rosetta refinement yielded a model with density correlation as good or better than any of the control methods for all but six structures. The dependence of success on sequence identity over the combined set is illustrated in Fig. 2b. The improvement in performance is particularly striking below 22% sequence identity, where the quality of the starting homology models becomes too low for the control methods in almost all cases. With the new method the success rate in the 15–28% sequence identity range, generally considered very challenging for molecular replacement, is over 50%.

Figure 2c illustrates the dependence of model-building on the quality of initial electron density. Conventional chain rebuilding requires a map in which the connectivity is largely correct (leftmost panel), whereas the new method can tolerate breaks in the chain more than other methods (panels 2–4), as long as there is sufficient information in the electron density map, combined with the Rosetta energy function, to guide structure optimization. The map on the far right contains too little information to guide energy-based structure optimization and hence the new approach fails. In the five blind cases that have not yet been solved the comparative models may have been too low in quality, or there may have been complications in the X-ray diffraction data sets themselves.

c, Dependence of structure determination success on initial map quality. Sigma-A-weighted $2mF_o - DF_c$ density maps (contoured at 1.5σ) computed from benchmark set templates with divergence from the native structure increasing from left to right are shown in grey; the solved crystal structure is shown in yellow. The correlation with the native density is shown above each panel. The solid green bar indicates structures the new approach was able to solve ($R_{\text{free}} < 0.4$); the red bar those that torsion-space refinement or DEN refinement is able to solve, and the purple bar those that can be solved directly using the template.

Key to the success of the approach described here is the integration of structure prediction and crystallographic chain tracing and refinement methods. Simulated annealing guided by molecular force fields and diffraction data has had an important role in crystallographic refinement^{14,21}. Structure prediction methods such as Rosetta can be even more powerful when combined with crystallographic data because the force fields incorporate additional contributions such as solvation energy and hydrogen bonding, and the sampling algorithms can build non-modelled portions of the molecule *de novo* and cover a larger region of conformational space than simulated annealing. The difference between Rosetta sampling and simulated annealing sampling, both using crystallographic data, is illustrated in Fig. 3. Beginning with the homology model placed by molecular replacement in the unit cell for blind case 6, we generated 100 models by simulated annealing at two starting temperatures, and 100 models with Rosetta energy- and density-guided optimization followed by refinement. The $2mF_o - DF_c$ (ref. 22) electron density maps generated using phases from over 50% of the Rosetta models had correlations 0.36 or better to the final refined map, whereas fewer than 5% of models from simulated annealing had correlations this high. Our approach probably outperforms even extreme simulated annealing because the physical chemistry and protein structural information which guide sampling eliminate the vast majority of non-physical conformations.

Approaches to molecular replacement combining the power of crystallographic map interpretation and structure prediction methodology

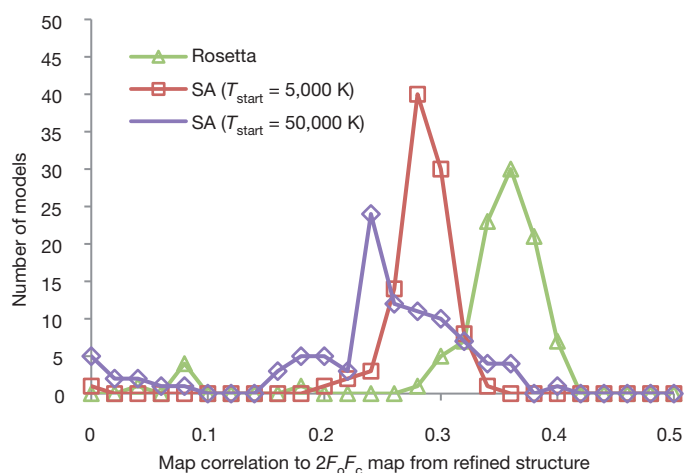


Figure 3 | Comparison of the effectiveness of model diversification using Rosetta and simulated annealing. For blind case 6, 100 models were generated using either simulated annealing with a start temperature of 5,000 K, simulated annealing with a start temperature of 50,000 K, or Rosetta energy- and density-guided optimization. The correlation between $2mF_o - DF_c$ density maps computed from each structure and the final refined density was then computed; the starting model has a correlation of 0.29 and the distributions of the refined models are shown in the figure. Rosetta models have correlations better than the initial model much more often than simulated annealing.

are likely to become increasingly useful in the next few years. First, the number of already-determined structures will continue increasing, making it increasingly likely that there will be a structure with the required >20% sequence identity: the chance there is a structure with a sequence identity of 20% or greater is more than twice that of finding a structure with at least 30% sequence identity²³. Second, as more work focuses on proteins that cannot be expressed in *Escherichia coli*, the currently preferred methods for experimental phase determination based on selenomethionine replacement may be more difficult to apply. Finally, as protein structure modelling algorithms improve, better initial models should further increase the radius of convergence of the approach.

METHODS SUMMARY

Starting models (templates) for molecular replacement were generated by searching the PDB using HHpred¹⁸ for proteins likely to have structures related to the query. Starting models were constructed from alignments generated by HHpred. Unaligned residues were removed from the template and non-identical side chains were stripped back to the gamma carbon (CG), as suggested in previous work⁹. An initial Phaser search with a low rotation function cutoff (50%) and modest packing threshold (up to 10 clashes) was used to find up to five putative molecular replacement (MR) solutions for each template. Each MR solution for each template was used to obtain an initial estimate of phases and the corresponding sigma-A-weighted $2mF_o - DF_c$ density map was generated²². Gaps in the initial alignment, as well as regions around deletions, were rebuilt using the Rosetta loop modelling protocol¹², which alternates insertion of short fragments with similar local sequences and cyclic coordinate descent (CCD) closure²⁴. Twenty-four rounds of side chain rotamer optimization and side chain and backbone torsion-space minimization were then used to optimize a linear combination of the Rosetta all-atom energy and a term assessing agreement to the electron density. Following the energy- and density-guided refinement, models were ranked based on the Phaser log-likelihood score. The highest ranked models were then subjected to a second round of modelling using the Rosetta iterative rebuild and refine protocol¹² constrained by density. After this final round of refinement, the model with best agreement to the experimental data (highest likelihood) was used to either find additional models in the asymmetric unit, or as a starting point for Phenix AutoBuild.

The procedures described here require considerable computation as up to several thousand Rosetta models are generated for each structure, typically requiring 0.5–1 h per structure of CPU time. We have developed automated procedures in Phenix (*phenix.mr_rosetta*) that use Rosetta and Phenix modules to carry out and extend many of the methods described here with density modification and density averaging, potentially allowing fewer Rosetta models to be used. All the methods described in this paper are available in release 3.2 of Rosetta.

Full Methods and any associated references are available in the online version of the alignment at www.nature.com/nature.

Received 4 August 2010; accepted 22 February 2011.

Published online 1 May 2011.

- Rossmann, M. G. *The Molecular Replacement Method* (Gordon & Breach, 1972).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
- Brünger, A. T. et al. Crystallography & NMR system: a new software system for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025 (1997).
- Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature Protocols* **3**, 1171–1179 (2008).
- Terwilliger, T. C. et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D* **64**, 61–69 (2008).
- DePristo, M. A., de Bakker, P. I. W., Johnson, R. J. K. & Blundell, T. L. Crystallographic refinement by knowledge-based exploration of complex energy landscapes. *Structure* **13**, 1311–1319 (2005).
- Cowan, K. The Buccaneer software for automated model building. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
- Schwarzenbacher, R., Godzik, A., Grzechnik, S. K. & Jaroszewski, L. The importance of alignment accuracy for molecular replacement. *Acta Crystallogr. D* **60**, 1229–1236 (2004).
- Rodríguez, D. D. et al. Crystallographic *ab initio* protein structure solution below atomic resolution. *Nature Methods* **6**, 651–653 (2009).
- Suhre, K. & Sanejouand, Y. H. On the potential of normal-mode analysis for solving difficult molecular-replacement problems. *Acta Crystallogr. D* **60**, 796–799 (2004).
- Qian, B. et al. High-resolution structure prediction and the crystallographic phase problem. *Nature* **450**, 259–264 (2007).
- Schröder, G., Levitt, M. & Brünger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464**, 1218–1222 (2010).
- Brünger, A. T., Kuriyan, J. & Karplus, M. Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458–460 (1987).
- Raman, S. et al. NMR structure determination for larger proteins using backbone-only data. *Science* **327**, 1014–1018 (2010).
- Das, R. & Baker, D. Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.* **77**, 363–382 (2008).
- Brünger, A. T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475 (1992).
- Söding, J. Protein homology detection by HMM–HMM comparison. *Bioinformatics* **21**, 951–960 (2005).
- Schröder, G. F., Brünger, A. T. & Levitt, M. Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. *Structure* **15**, 1630–1641 (2007).
- Brünger, A. T. Extension of molecular replacement: a new search strategy based on Patterson correlation refinement. *Acta Crystallogr. A* **46**, 46–57 (1990).
- Brünger, A. T., Karplus, M. & Petsko, G. A. Crystallographic refinement by simulated annealing: application to crambin. *Acta Crystallogr. A* **45**, 50–61 (1989).
- Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42**, 140–149 (1986).
- Vitkup, D., Melamud, E., Moul, J. & Sander, C. Completeness in structural genomics. *Nature Struct. Biol.* **8**, 559–566 (2001).
- Canutescu, A. & Dunbrack, R. Cyclic coordinate descent: a new algorithm for loop closure in protein modeling. *Protein Sci.* **12**, 963–972 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements R.J.R., T.C.T. and D.B. thank the NIH (5R01GM092802), the Wellcome Trust (R.J.R.), and HHMI (D.B.) for funding this research. F.D. acknowledges the NIH (P41RR002250) and HHMI. D.F. and A.A. acknowledge support from the Israel Science Foundation. G.O. thanks DK Molecular Enzymology (FWF-project W901) and the Austrian Science Fund (FWF-project P19858). The work of A.W. was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. H.I. acknowledges support from the academy of Finland (1131413). S.M.V. was supported by a grant from the Protein Structure Initiative of National Institute of General Medical Sciences (U54 GM074958). The work of P.R.P. at Argonne National Laboratory was supported by the US Department of Energy's Office of Science, Biological and Environmental Research GTL programme under contract DE-AC02-06CH11357. We thank all members of the JCSG for their general contributions to the protein production and structural work. The JCSG is supported by the NIH, National Institutes of General Medical Sciences, Protein Structure Initiative (U54 GM094586 and GM074898).

Author Contributions F.D., T.C.T., R.J.R. and D.B. developed the methods described in the manuscript; F.D., T.C.T., R.J.R., A.W. and D.B. wrote the paper. A.W., G.O., U.W., E.V., A.A., D.F., H.L.A., D.D., S.M.V., H.I. and P.R.P. provided the data and refined one or more structures to completion.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.C.T. (terwilliger@lanl.gov) or D.B. (dabaker@u.washington.edu).

METHODS

Preparation of templates and identification of initial molecular replacement solutions. For the application of the new method to blind cases, templates were identified using HHpred¹⁸. For both the blind and benchmark data sets, HHpred was used to generate initial alignments. We prepared templates by removing all unaligned residues and stripping all non-identical side chains to the gamma carbon (CG), as suggested in previous work⁹. An initial Phaser search with a low rotation function cutoff (50%) and modest packing threshold (up to 10 clashes) was used to find up to five putative MR solutions for each template. In two blind cases (12 and 13 in Table 1), Phaser was unable to locate the correct configuration of a molecule using the template alone, but modelling in Rosetta without density-fitting constraints before Phaser search enabled discovery of the correct rigid-body placement of the molecule, with very low Phaser translation function Z-scores (TFZ) of 4–6 (after solving 13, it was discovered that breaking the template into two rigid subunits enabled solution of the molecular replacement problem). If point-group symmetry was present in the templates, the initial search (and subsequent steps) were carried out both with monomeric and multimeric models (see subsection on symmetric modelling into density below).

Rebuilding and refinement into density. Each MR solution for each template was used to obtain an initial estimate of phases and the corresponding sigma-A-weighted $2mF_o - DF_c$ density map was generated²². Gaps in the initial alignment, as well as regions around deletions, were rebuilt using the Rosetta loop modelling protocol¹², which alternates insertion of short fragments with similar local sequences and CCD closure²⁴. Twenty-four rounds of side chain rotamer optimization and side chain and backbone torsion-space minimization were then used to optimize a linear combination of the Rosetta all-atom energy and a term assessing agreement to the electron density. Agreement to density was computed using an extension of a method previously developed for building into cryo-electron microscopy density²⁵. Density was calculated from a model using a single-Gaussian approximation to atomic scattering factors. Correlation coefficients between model and map were calculated for each residue: the computed density includes all atoms in the residue and the backbone in the two flanking residues on each side, and the correlation is taken over a mask extending 5 Å from each atom. Scores are proportional to the negative log probability that observed correlations occur by random chance, assuming a normal distribution; parameters are trained matching randomly oriented fragments into synthesized density. In all cases, density was truncated at 3 Å.

Following the energy- and density-guided refinement, models were ranked based on the Phaser log-likelihood score. The highest ranked models were then subjected to a second round of modelling using the Rosetta iterative rebuild and refine protocol¹² constrained by density. Regions that deviated the most from the current estimate of the electron density were rebuilt; clashes between crystallographic (and non-crystallographic) contacts were also always rebuilt. For each template carried over to the second round (typically the top-scoring 3–10 models from the previous round), 2,000 Rosetta models were generated. The likelihood of the diffraction data was again computed using Phaser for the lowest-energy 10% of models, and if the five highest likelihood models were in the same rigid-body configuration (that is if they had density correlations above 0.2 with each other), they were used to re-phase the density and an additional round (24 cycles) of side chain optimization and refinement was carried out in Rosetta. If the top-scoring models differed, then additional templates were considered (if available) or Rosetta homology modelling was used to perturb the initial structures before molecular replacement.

After this final round of refinement, the model with best agreement to the experimental data (highest likelihood) was used to either find additional models in the asymmetric unit, or as a starting point for Phenix AutoBuild. In cases where the R_{free} was better than random but higher than 0.4, and a majority of residues were placed, additional refinement was carried out using models produced by AutoBuild, which allows for recovery from sequence alignment errors. The bond lengths and bond angles were first replaced with ideal values with small compensating changes in the torsion angles to minimize the change in interatomic distances, and the idealized models were then subjected to 48 cycles of side chain rotamer optimization and side chain and backbone torsion minimization. In the first 24 cycles, the Rosetta all-atom energy function was optimized, and in the final 24 cycles a weighted sum of Rosetta all-atom energy and the fit-to-density energy described above was optimized.

Refinement of symmetric complexes into density. Key to solving many of the blind cases was proper treatment of symmetry. In cases where there is point-group symmetry in the asymmetric unit (either from the template or subsequently discovered by molecular replacement search) or there is close contact between crystal partners, the Rosetta symmetric modelling framework²⁶ was used to reduce the size of the conformational space which must be searched. This occurred in blind

cases where either there was point-group symmetry in the template(s) (6 in Table 1), point-group symmetry was found during the Phaser search (13), or tight crystal contacts formed point-group symmetry (8 and 10). In these cases, Rosetta optimizes only the torsion angles in one subunit and the rigid-body degrees of freedom of the corresponding symmetric group. The energy is calculated explicitly over a non-redundant subset of atoms for computational efficiency, but the fit to density is calculated without symmetrization. This is similar to the “strict formulation” of symmetry introduced in ref. 27.

Symmetric modelling in Rosetta requires that the energy of a symmetric complex be expressible in terms of a single subunit or as pairwise interactions between this subunit and other ones. Minimization also only considers gradients from these components. To take advantage of Rosetta's symmetric modelling with asymmetric density data, the gradients of each subunit with respect to the fit-to-density energy must be mapped to a single subunit. The score of a residue i 's fit to density is just the sum of the fit-to-density scores over all of i 's copies. As a first approximation, the gradient at i can be computed as the combined gradients of all of i 's copies, rotated by the symmetry operation to rotate the subunit containing i 's copy to the one containing i . Unfortunately, although this approach correctly handles gradients of internal torsions, the gradients at each symmetric degree-of-freedom are not correctly handled. Proper handling takes advantage of the formulation from ref. 28 to efficiently convert Cartesian gradients to torsion-space gradients. For each atom in the symmetric complex, we compute F_1 and F_2 corresponding to the unrotated gradient with respect to the fit-to-density score. For internal torsional degrees of freedom, the rotation applied to each F_1/F_2 just maps each subunit back to the asymmetric unit. At each symmetric degree of freedom we apply a corresponding symmetry operation; for example, in D3 symmetry (a dimer of trimers) the degree of freedom corresponding to the “spin” of the trimers applies the rotation used to transform between trimers to all the F_1/F_2 's in one of the trimers.

Refinement against the Patterson function. In benchmark cases where the Phaser translation search failed to find the correct molecular placement even when many potential solutions were considered, we conducted refinement against the Patterson function. A score function was implemented that assessed the correlation between the computed and experimental Patterson map (next paragraph). The map was truncated to between 3.5 Å and 10 Å resolution (in reciprocal space) and 5 Å to ~75% of the template diameter (in real space). Starting models used the same templates and rebuilding procedure as the density refinement. Because the correct rotation is not known at this stage, the molecule orientation was randomized at the beginning of each refinement trajectory and constraints on backbone atoms were used to prevent the molecule from rotating more than ~5° from this starting orientation.

The scoring function we optimize is the weighted sum of Rosetta's all-atom potential function and the correlation between the calculated Patterson map and the observed Patterson map. To make this tractable in Rosetta refinement, which may require tens of thousands of score-function evaluations per trajectory, simplifications are necessary. Directly computing $\partial p_{\text{calc}}/\partial x$ requires three fast Fourier transforms (FFTs) per atom. However, since what is needed is not $\partial p_{\text{calc}}/\partial x$ but instead the sum $\partial \sum_{\text{map}} p_{\text{calc}} p_{\text{obs}}/\partial x$, FFTs can be used to compute the change in correlation at every position in the map at once (where p is the Patterson density and ρ is the real-space density):

$$\frac{\partial \sum_{\text{map}} p_{\text{calc}} p_{\text{obs}}/\partial x}{\partial x} = F^{-1}[F[p_{\text{obs}}] \cdot F[\rho_{\text{calc}}] \cdot [F[\partial \rho_{\text{calc}}/\partial x]]](x) \quad (1)$$

Assuming a fixed B-factor over the molecule, this requires just 3 FFTs per atom type (the correction terms that make this not just the overlap integral but a true correlation can be folded into the same FFT). Then, given a model to refine against the Patterson map, we compute equation (1) once, sum over all the symmetric orientations of the space group, and interpolate the gradient at each atom's position. Given sufficiently fine sampling, this gives a very close approximation to the true derivative in a small fraction of the CPU time.

For side chain optimization, where we must rescore the Patterson correlation for exponentially many combinations of side chain rotamers, exact computation is also intractable. However, first computing the density ρ_{calc} of the backbone only, then computing the correlation scores for each side chain rotamer independently, provides a reasonably good approximation with only several hundred to several thousand function evaluations (one for each rotamer).

Torsion space simulated annealing with DEN restraints. As a control, we ran torsion-space simulated annealing with DEN restraints¹³ on the blind tests and on the complete benchmark set of structures related to PDB entries 1XVQ and 1A2B. Using the same template and placement used by Rosetta refinement, initial homology models were built in Modeller²⁹ (using the same alignment used by Rosetta). DEN refinements were carried out using the `refine_lowres.inp` script distributed

with CNS version 1.3 as a template. The results of these analyses for the benchmark set of structures are shown in Supplementary Tables 4 and 5, and for the blind tests, as part of Table 1.

Massive-sampling simulated annealing. To test the role that massive sampling around the conformation of the input structure plays in the success of our new methods, we developed an 'extreme simulated annealing protocol', where 1,000 models were produced by simulated annealing refinement, the best of these models is used as the starting point for automated model rebuilding, density modification and refinement with PHENIX, and the resulting model is used as the starting point for a second iteration of the procedure. In this procedure, simulated annealing was carried out in phenix.refine using the flag 'simulated_annealing = True' and the default starting temperature of 5,000 K.

Implementation in Phenix and Rosetta. The procedures described here require considerable computation as up to several thousand Rosetta models are generated for each structure, typically requiring 0.5–1 h per structure of CPU time. We have developed automated procedures in Phenix (phenix.mr_rosetta) that use Rosetta and Phenix modules to carry out and extend many of the methods described here with density modification and density averaging, potentially allowing fewer Rosetta models to be used. Beginning with correctly placed templates (including all copies of each molecule, and placed domains for 13), each of 13 blind test cases in Table 1 can be solved with phenix.mr_rosetta using 20 Rosetta models during each rebuilding cycle, yielding free *R* values of 0.42 or lower (mean *R*_{free} = 0.33), and requiring from approximately 30 to 130 CPU-hours to complete.

All the methods described in this paper are available in release 3.2 of Rosetta. An application, 'mr_protocols,' is included which was used (together with Phaser and Phenix Autobuild) to generate all the results in this paper. The flags files used for Rosetta are shown below.

Comparative modelling (with target sequence target.fasta, alignment target_template.ali, and template template.pdb) in the context of density:

```
-database $DB
-MR:mode cm
-in:file:extended_pose 1
-in:file:fasta target.fasta
-in:file:alignment target_template.ali
-in:file:template_pdb template.pdb
-loops:frag_sizes 9 3 1
-loops:frag_files aa1xxx_09_05.200_v1_3.gz aa1xxx_03_05.200_v1_3.gz none
-loops:random_order
-loops:random_grow_loops_by 5
-loops:extended
-loops:rebuild quick_ccd
-loops:relax relax
-relax:default_repeats 4
-relax:jump_move true
-edensity:mapreso 3.0
-edensity:grid_spacing 1.5
-edensity:mapfile target.map
-edensity:sliding_window_wt 1.0
-edensity:sliding_window 5
-cm:aln_format grishin
-MR:max_gaplength_to_model 10
-nstruct $STRUCTS
```

In cases where Rosetta was used to 'pre-refine' the structure before Phaser, the same command line was used without the -edensity:* flags. Modelling with symmetry used the flags above in addition to the flag '-symmetry_definition symm.def', where symm.def defines the symmetry in the template. Symmetry definition file creation is automated using a script; see the Rosetta documentation for more details.

Additional refinement (both after comparative modelling and after autobuilding in some cases):

```
-database $DB
-MR:mode relax
-in:file:rosetta_model.pdb
-relax:default_repeats 4
-relax:jump_move true
-edensity:mapreso 3.0
-edensity:grid_spacing 1.5
-edensity:mapfile target.map
-edensity:sliding_window_wt 1.0
-edensity:sliding_window 5
-nstruct 5
```

Comparative modelling against the Patterson function (the experimental Patterson map, target_pat.map, is computed outside Rosetta):

```
-MR:mode cm
-in:file:extended_pose 1
-in:file:fasta target.fasta
-in:file:alignment target_template.ali
-in:file:template_pdb template.pdb
-loops:frag_sizes 9 3 1
-loops:frag_files aa1xxx_09_05.200_v1_3.gz aa1xxx_03_05.200_v1_3.gz none
-loops:random_order
-loops:random_grow_loops_by 5
-loops:extended
-loops:rebuild quick_ccd
-loops:relax relax
-relax:default_repeats 2
-relax:jump_move true
-edensity:grid_spacing 1.6
-edensity:mapfile target_pat.map
-edensity:use_spline_interpolation true
-edensity:realign random
-edensity:use_symm_in_pcalc true
-edensity:patterson_lowres_limit 3.5
-edensity:patterson_hires_limit 10.0
-edensity:patterson_minR 5.0
-edensity:patterson_maxR 14.0
-edensity:patterson_B 0.2
-edensity:patterson_cc_wt 0.5
-cm:loop_rebuild_filter 500
-cm:aln_format grishin
-cm:max_loop_rebuild 10
-cm:min_loop_size 4
-MR:max_gaplength_to_model 10
-nstruct $STRUCTS
```

Most of the data used in this paper is available at http://www.phenix-online.org/phenix_data/terwilliger/rosetta_2011/ (additional blind cases will be made available as the structures are deposited).

25. DiMaio, F., Tyka, M. D., Baker, M. L., Chiu, W. & Baker, D. Refinement of protein structures into low-resolution density maps using Rosetta. *J. Mol. Biol.* **392**, 181–190 (2009).
26. André, I., Bradley, P., Wang, C. & Baker, D. Prediction of the structure of symmetrical protein assemblies. *Proc. Natl Acad. Sci. USA* **104**, 17656–17661 (2007).
27. Weiss, W. I., Brünger, A. T., Skehel, J. J. & Wiley, D. D. Refinement of the influenza virus hemagglutinin by simulated annealing. *J. Mol. Biol.* **212**, 737–761 (1990).
28. Abe, H., Braun, W., Noguti, T. & Gö, N. Rapid calculation of first and second derivatives of conformational energy with respect to dihedral angles for proteins general recurrent equations. *Comput. Chem.* **8**, 239–247 (1984).
29. Eswar, N. *et al.* Comparative protein structure modeling with MODELLER. *Curr. Protoc. Bioinform. (Suppl.)* **15**, 5.6 doi:10.1002/0471250953.bi0506s15 (2006).

UBCH7 reactivity profile reveals parkin and HHARI to be RING/HECT hybrids

Dawn M. Wenzel¹, Alexei Lissounov¹, Peter S. Brzovic¹ & Rachel E. Klevit¹

Although the functional interaction between ubiquitin-conjugating enzymes (E2s) and ubiquitin ligases (E3s) is essential in ubiquitin (Ub) signalling, the criteria that define an active E2–E3 pair are not well established. The human E2 UBCH7 (also known as UBE2L3) shows broad specificity for HECT-type E3s¹, but often fails to function with RING E3s *in vitro* despite forming specific complexes^{2–4}. Structural comparisons of inactive UBCH7–RING complexes with active UBCH5–RING complexes reveal no defining differences^{3,4}, highlighting a gap in our understanding of Ub transfer. Here we show that, unlike many E2s that transfer Ub with RINGs, UBCH7 lacks intrinsic, E3-independent reactivity with lysine, explaining its preference for HECTs. Despite lacking lysine reactivity, UBCH7 exhibits activity with the RING-in-between-RING (RBR) family of E3s that includes parkin (also known as PARK2) and human homologue of ariadne (HHARI; also known as ARIH1)^{5,6}. Found in all eukaryotes⁷, RBRs regulate processes such as translation⁸ and immune signalling⁹. RBRs contain a canonical C3HC4-type RING, followed by two conserved Cys/His-rich Zn²⁺-binding domains, in-between-RING (IBR) and RING2 domains, which together define this E3 family⁷. We show that RBRs function like RING/HECT hybrids: they bind E2s via a RING domain, but transfer Ub through an obligate thioester-linked Ub (denoted ~Ub), requiring a conserved cysteine residue in RING2. Our results define the functional cadre of E3s for UBCH7, an E2 involved in cell proliferation¹⁰ and immune function¹¹, and indicate a novel mechanism for an entire class of E3s.

RING and U-box E3s facilitate Ub transfer directly from an activated E2~Ub to a lysine on a target protein. Therefore, E2s that function with RINGs must be catalytically competent to form an isopeptide bond between Ub and lysine. Previous characterization of E2 activity demonstrates that some E2s can transfer Ub to free lysine independent of an E3 (ref. 12), providing a framework to examine E2 function. We compared the intrinsic, E3-independent reactivity of UBCH7~Ub and UBCH5C~Ub with free amino acids that represent Ub acceptors: lysine, serine, threonine¹³, cysteine¹⁴, or arginine as a control (Fig. 1a). UBCH5C~Ub reacts completely with either cysteine or lysine, but not other amino acids, indicating the side-chain

functional group is the relevant nucleophile. Notably, UBCH7 reacts only with cysteine. Reaction time courses for UBCH7~Ub and UBCH5C~Ub with free lysine show that UBCH5C~Ub is nearly depleted after 15 min, while after 60 min UBCH7~Ub shows no detectable reaction (Fig. 1b). The lack of reactivity of UBCH7 is lysine-specific and cannot be attributed to UBCH7~Ub being intrinsically more stable, as both E2s react equally rapidly with cysteine (Supplementary Fig. 2a). UBE2K and UBC13, E2s known to function with RINGs^{15,16}, both react with cysteine and lysine (Supplementary Fig. 2b), indicating that lysine reactivity is a general feature of RING-active E2s. The reactivity properties of UBCH7 are conserved, as the *Caenorhabditis elegans* orthologue Ubc18 (ref. 17) also lacks lysine reactivity (Supplementary Fig. 3).

To determine which residues in E2s are important for lysine reactivity, the active site sequences of lysine-reactive E2s were aligned with that of UBCH7 (Fig. 2a). Two residues in UBCH7 are distinctly different: D87 and D117 (in UBCH5C numbering) are proline and histidine, respectively, in UBCH7. To establish whether these residues contribute to lysine reactivity, each was mutated in UBCH5C and lysine reactivity was measured (Fig. 2b). The effect of substitution at position 87 ranges from no effect for the isosteric mutation D87N to complete loss of lysine reactivity for the charge-swapped D87K mutation. UBCH5C(D87E) and UBCH5C(D87P) have intermediate reactivities. Consistent with D87 having a general role in lysine reactivity, mutation of the analogous residue in UBE2K(D94E) results in decreased lysine reactivity and impaired formation of free poly-Ub chains (Supplementary Fig. 4a, b). Substitution of D117 in UBCH5C with a histidine as found in UBCH7 greatly decreases lysine reactivity (Fig. 2b). A structurally analogous residue in the SUMO E2 (D127 in UBC9), has been shown to lower the pK_a of a lysine approaching the active site¹⁸. The invariant active-site asparagine residue (N77 in UBCH5C) is recognized for its role in isopeptide catalysis¹⁹, and its mutation to serine abolishes UBCH5C~Ub lysine reactivity in our assay. Identification of several residues that affect the intrinsic lysine reactivity of E2s indicates that the determinants are probably multifactorial. Accordingly, we failed to convert UBCH7 into a lysine-reactive E2 by mutation (Supplementary Table).

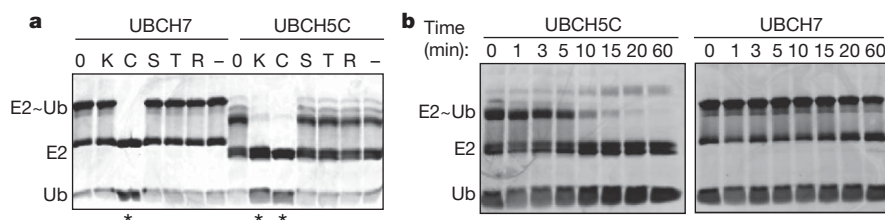


Figure 1 | UBCH7 does not react with free lysine. **a**, Coomassie-stained SDS-PAGE of UBCH7~Ub (left) and UBCH5C~Ub (right) incubated with amino acids lysine, serine, threonine, arginine, or buffer (-). Reactions were quenched in non-reducing loading buffer. Starting amounts of E2~Ub before amino acid addition are indicated as '0'. Reactivity with amino acids is indicated by loss of

E2~Ub and concurrent increase in free Ub (denoted by asterisks) and free E2. **b**, Time-course assays of UBCH5C~Ub and UBCH7~Ub incubated with lysine. Reactions were quenched in non-reducing loading buffer at the indicated times.

¹Department of Biochemistry, Box 357350, University of Washington, Seattle, Washington 98195, USA.

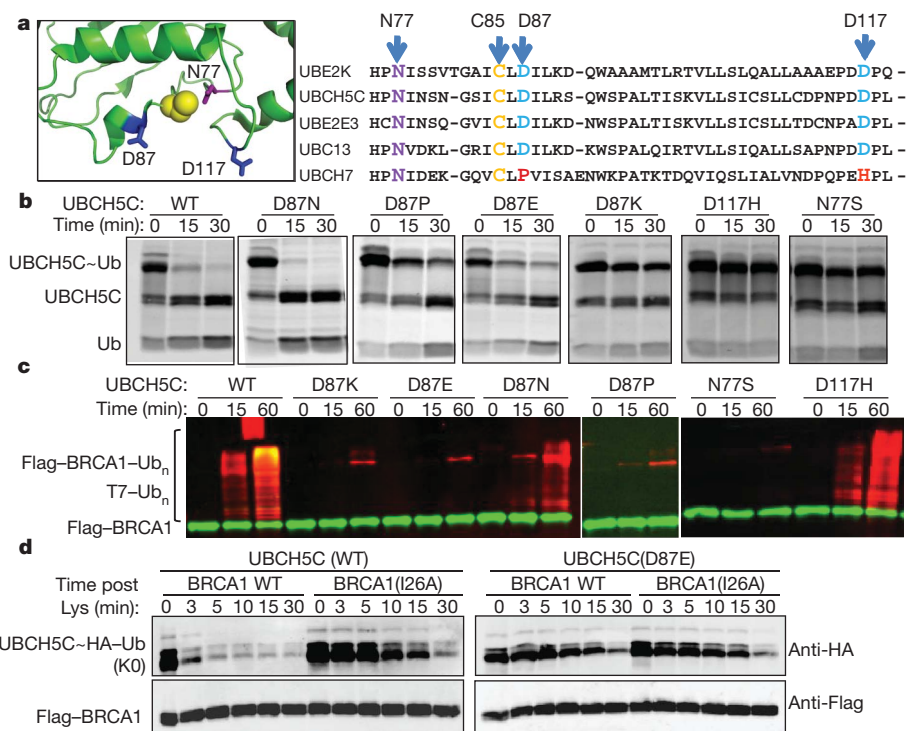


Figure 2 | Lysine reactivity is multifactorial. **a**, Alignment of E2 active-site residues (right). Structure of UBCH5C (green; PDB accession code 2FUH) with active-site residues represented as sticks and active-site cysteine as spheres (left). **b**, Wild-type (WT) or mutant UBCH5C~Ub incubated with lysine as in Fig. 1b. **c**, Western blot of Flag-BRCA1(1-304)/BARD1(26-327) (green) auto-ubiquitination with indicated E2 and T7-Ub (red). Time is measured as

minutes after ATP addition. Multiply ubiquitinated species are indicated as 'Ub_n'. **d**, Same as in Fig. 1b, except a 1:1 (E3:E2) equivalent of Flag-BRCA1(1-112)/BARD(26-140) was added to either wild-type UBCH5C (left) or the D87E mutant (right) charged with HA-Ub where all lysines have been mutated to arginines (indicated as 'K0'). Reactions were visualized by western blot for HA (Ub(K0)) and Flag (BRCA1) epitopes.

To assess whether the intrinsic reactivity of an E2 is predictive of its functional E3 interaction, UBCH5C mutants were assayed with RING and HECT-type E3s. Of the D87 mutants, only UBCH5C (D87N) (which has wild-type lysine reactivity) is able to function with the RING E3 BRCA1/BARD1 in an auto-ubiquitination assay; the other substitutions are inactive (Fig. 2c). The observed loss of activity results from E2 catalytic defects, as NMR binding experiments confirm that UBCH5C(D87) mutants bind BRCA1 comparably to wild-type UBCH5C (Supplementary Fig. 5). UBCH5C(D117H), which has impaired lysine reactivity, retains some ability to transfer Ub to BRCA1 (Fig. 2c). Given its position at the active site, D117 may provide substrate-specific lysine reactivity as a gating residue. As expected, UBCH5C(N77S) is inactive with BRCA1. Previous studies show that mutation of E2 residue N77 abolishes Ub transfer with RING-type ligases MDM2 (ref. 19), RMA1 (also known as RNF5) (ref. 20), CNOT4 (CCR4-NOT transcription complex subunit 4) and APC/C (anaphase promoting complex/cyclosome) (ref. 21) but retains activity with HECT-type ligases E6-AP (also known as UBE3A), KIAA10 (also known as UBE3C; ref. 19) and NEDD4L (ref. 22). This indicates that the catalytic requirements for trans-thiolation differ from those for isopeptide bond synthesis. Accordingly, the ability to form E3~Ub thioesters with the HECT E3 E6-AP is unaffected in all UBCH5C mutants (Supplementary Fig. 6). Taken together, our results indicate that E2 lysine reactivity is a prerequisite for transfer with RING E3s. Furthermore, lysine-unreactive E2s, such as UBCH5C(N77S), can be diagnostic for differentiating between RING- versus HECT-type Ub transfer mechanisms.

Although many E2s possess an intrinsic reactivity for lysine, most do not transfer Ub to protein substrates independent of an E3. To explore the contribution of E3s to E2 reactivity, we examined the intrinsic reactivity of E2~Ub in the presence of RING or HECT E3s. Interaction with BRCA1/BARD1 (RING) or E6-AP (HECT) does not change the intrinsic reactivity of UBCH5C and UBCH7 with free

amino acids (Supplementary Fig. 7a, b). However, BRCA1/BARD1 enhances UBCH5C~Ub lysine reactivity compared to a reaction containing the E2-binding mutant BRCA1(I26A)^{3,15} (Fig. 2d). Intriguingly, the lysine reactivity of UBCH5C(D87E), which retains both its ability to bind BRCA1/BARD1 and some intrinsic lysine reactivity, is not enhanced by BRCA1 (Fig. 2d). These results indicate that RING binding to a lysine-reactive E2~Ub results in a thioester with enhanced reactivity to lysine and residues such as D87 in UBCH5C couple E3 binding and E2 activation. Notably, E6-AP does not enhance UBCH5C~Ub lysine reactivity (Supplementary Fig. 8b), highlighting a mechanistic difference between RINGs and HECTs. In matched experiments with UBCH7, neither BRCA1/BARD1 nor E6-AP enhance reactivity towards free cysteine (Supplementary Fig. 8a, c).

Besides HECT-type ligases, UBCH7 is reported to function with members of the RING E3 family known as RBRs, which includes parkin and HHARI^{5,6}. This activity runs contrary to our conclusion that UBCH7 lacks lysine reactivity and is consequently restricted to Ub transfer involving trans-thiolation chemistry. Therefore, we examined the Ub ligase mechanism for HHARI and parkin. The minimal ligase (RBR) domains of HHARI_{R1-IBR-R2} (where R1 and R2 are RING1 and 2, respectively) and parkin_{R1-IBR-R2} (see Supplementary Fig. 9 for schematic of constructs) show comparable auto-ubiquitination activity with either UBCH7 or UBCH5C (Fig. 3a, b). In contrast to our results with BRCA1/BARD1, UBCH5C(N77S) exhibits Ub transfer with HHARI_{R1-IBR-R2} and parkin_{R1-IBR-R2} (Fig. 3a, b). As UBCH5C(N77S) activity is restricted to HECT-type ligases, its activity with parkin and HHARI indicates that these ligases do not function via a typical RING-type mechanism.

A hallmark of HECT-type Ub transfer is the formation of an obligate E3~Ub thioester intermediate. Reactions designed to trap a HHARI~Ub conjugate were conducted on ice and quenched with SDS-loading buffer with or without reducing agent (βME) to distinguish reducible thioester-linked Ub from non-reducible isopeptide-linked Ub. A βME-sensitive band corresponding to the molecular

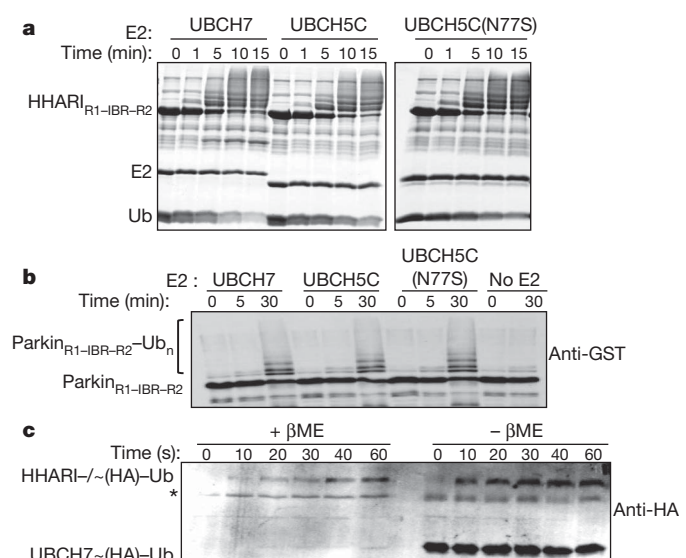


Figure 3 | RBRE E3s function via a HECT-like mechanism. **a**, HHARI_{R1-IBR-R2} auto-ubiquitination assays with the indicated E2 were visualized on a Coomassie-stained reducing SDS gel. Time is measured as minutes after ATP addition. **b**, Auto-ubiquitination assay of parkin_{R1-IBR-R2} with the E2s indicated. Products were visualized by western blotting for the GST tag on parkin. **c**, UBCH7 was pre-charged with HA-Ub and mixed with HHARI_{R1-IBR-R2}. Reactions were quenched in SDS buffer under reducing (+βME) and non-reducing conditions (−βME) and visualized by western blotting for the HA epitope on Ub. A βME-sensitive HA-Ub band corresponding to the molecular weight of HHARI_{R1-IBR-R2}-Ub appears at 10, 20 and 30 s after addition of UBCH7~Ub. Asterisk denotes a cross-reactive band.

weight of HHARI~Ub was detected at 10, 20, and 30 s after addition of pre-charged UBCH7~Ub (Fig. 3c). We next sought to determine the position of the HHARI active-site cysteine. The N-terminal canonical RING1 of HHARI has been shown to be the principal E2-binding region²³, although RING2 is also required for ligase activity⁵. Cysteine 357 in RING2 is highly conserved across RBR ligases (Fig. 4a). C357 is not a Zn²⁺-liganding residue, and mutation of C357 does not destabilize the RING2 structure⁵ (Supplementary Fig. 10), but does abolish the ability of HHARI to transfer Ub⁵. This indicates that C357 may have a catalytic function. Although HHARI_{R1-IBR-R2}(C357A) showed no ligase activity in an auto-ubiquitination assay, HHARI_{R1-IBR-R2}(C357S) generated a single

monoubiquitinated species (Fig. 4b). An analogous oxyester-linked Ub product can be generated on E2s for which the active-site cysteines have been mutated to serines^{20,22,24}. Consistent with its identity as an oxyester (versus isopeptide) bond, the Ub adduct on HHARI_{R1-IBR-R2}(C357S) is labile to alkaline treatment (Supplementary Fig. 11). Formation of an oxyester is unique to C357 as serine substitution of other conserved cysteines (both Zn²⁺- and non-Zn²⁺-liganding) in RING2 does not stall the E3 at a single-Ub adduct but rather impairs (C375S) or abolishes (C367S, C372S) ligase activity (Supplementary Fig. 12). Similarly, the parkin mutant C431S (analogous to HHARI(C357)) eliminates E3 ligase activity (Fig. 4c). We were unable to trap a parkin_{R1-IBR-R2}(C431S) Ub adduct under our reaction conditions. However, we note that the parkin mutation C431F has been consistently shown to abolish parkin's ubiquitination of substrates^{25–27} and genetically predisposes for Parkinson's disease²⁸. The results presented above are not affected by the presence of a GST domain, as several results were reproduced with non-GST versions of HHARI (Supplementary Fig. 13).

HHARI_{R1-IBR-R2}(C357A) effectively binds E2 as GST pull-downs with purified UBCH7 and constructs of GST-HHARI demonstrate that HHARI_{R1-IBR-R2}(C357A) interacts with UBCH7 as efficiently as wild-type HHARI_{R1-IBR-R2} (Fig. 4d). In contrast, a HHARI-RING1 mutant (I188A²³, analogous to BRCA1(I26A)) does not interact detectably with UBCH7 in this assay. Furthermore, C357 is surface accessible and reactive, as wild-type HHARI_{R2} but not HHARI_{R2}(C357A) is readily derivitized by cysteine-modifying reagents (Supplementary Fig. 14).

In the absence of a bona fide substrate, an *in vitro* product of HHARI-catalysed Ub transfer is the non-reducible ubiquitination of UBCH7 (Fig. 4b). Ubiquitination of UBCH7 is E3-dependent, and mutation of HHARI_{R1-IBR-R2}(C357) to serine or alanine abolishes the formation of this product (Fig. 4b). Our finding that HHARI_{R1-IBR-R2}(C357S) forms an oxyester-linked Ub without subsequent transfer of Ub to UBCH7 indicates an ordered mechanism that involves formation of an E3~Ub before modification of UBCH7 (or substrates). Thus, HHARI, unlike other RING E3s, does not facilitate direct transfer from an E2~Ub to a target. We note that the HHARI oxyester-linked Ub conjugate accumulates in low yield compared to the available number of active sites. This is consistent with our failure to observe intrinsic serine reactivity for UBCH7, even at serine concentrations as high as 0.5 M (Fig. 1a and data not shown) and suggests that the unique chemical environment surrounding a target residue (in this case an enzyme active site) contributes to catalysis—a contribution that is absent in the nucleophile

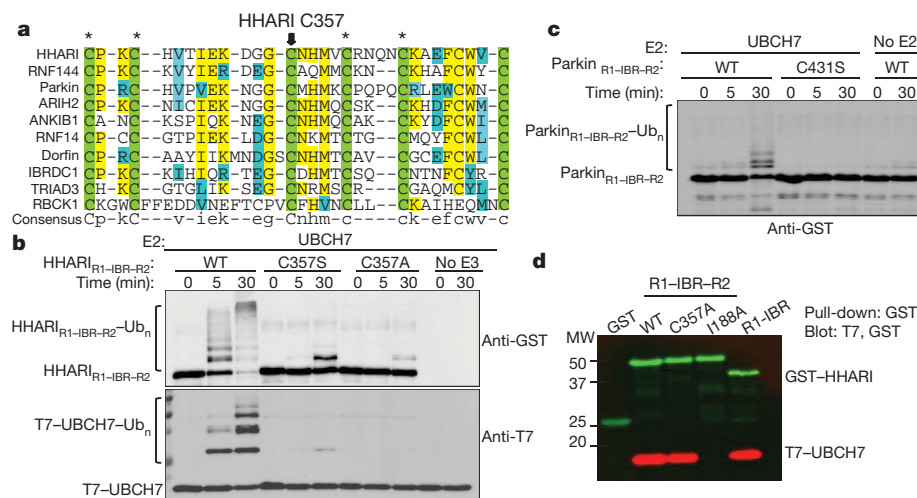


Figure 4 | Cysteine C357 is the active site of HHARI. **a**, Sequence alignment of RING2 domains from human RBR ligases. Conserved cysteine residues are coloured green. Asterisks indicate Zn²⁺-liganding cysteines in the HHARI_{R2} structure⁵. **b**, Auto-ubiquitination assays of T7-tagged UBCH7 and the indicated construct of HHARI. Ubiquitination is measured as time after ATP addition. Reaction products were visualized by western blotting simultaneously for the GST tag on HHARI (top panel) and the T7 tag on UBCH7 (bottom panel). **c**, E3 auto-ubiquitination assay with UBCH7 and the indicated construct of parkin. **d**, GST pull-downs of purified constructs of GST-tagged HHARI and T7-tagged UBCH7. Bound protein was eluted and visualized by western blotting for T7 (red) and GST (green) epitopes simultaneously.

assay. Our combined results are consistent with HHARI and parkin functioning via a HECT-like mechanism whereby RING1 harbours the E2-binding site and RING2 harbours the active-site cysteine.

Our characterization of the reactivity of UBCH7 resulted in two unexpected discoveries. First, whereas E2s known to work with RING-type E3s have E3-independent reactivity towards lysine, the intrinsic reactivity of UBCH7 is restricted to cysteine and, consequently, its Ub transfer activity is restricted to HECT-type ligases. Second, in the process of confirming reports that UBCH7 is active with E3s in the RBR family, we discovered that these RING-containing E3s function like HECTs in that they require an obligate trans-thiolation step during Ub transfer. Both findings have important implications for guiding our understanding of ubiquitination pathways. RING1 of HHARI does not harbour catalytic activity, and neither HHARI_{R1-IBR} nor HHARI_{R2} enhance lysine reactivity of UBCH5C (Supplementary Fig. 15). Our results underscore the diversity of structures that facilitate thiol-based Ub transfer, enzymes that include bacterial HECT-like E3s that bear no homology to eukaryotic HECT counterparts²⁹. Knockdown and over-expression studies indicate that UBCH7 regulates S-phase progression into G2, but neither the E3(s) nor targets responsible have been identified¹⁰. Our results indicate that the relevant E3(s) will be found among HECT or RBRs. Although it is possible that UBCH7 cooperates with RINGs such as BRCA1 to modify substrate cysteines, we have not observed such species.

Among human E2s, only five residue types are found at the position analogous to D87 of UBCH5C: aspartate, serine, asparagine, glutamate and proline. The tolerance for asparagine and serine at position 87 indicates that the negative charge of D87 may not be critical for its role in catalysis, but instead a hydrogen-bonding function seems likely, possibly interacting with the conjugated Ub. Of human E2s, only UBCH7 and UBCH8 have a proline at position 87. Although we did not test UBCH8 for lysine reactivity, it functions primarily with the HECT-type ligase HERC5 to transfer the Ub-like protein ISG15 to substrates³⁰. Like UBCH7, we anticipate UBCH8 activity will be limited to HECT or RBR E3s. Our effort to understand the mechanism of UBCH7-mediated Ub transfer highlights the predictive power of elucidating E2 mechanisms to understand the E3s with which they function.

METHODS SUMMARY

Plasmids, protein expression/purification, E3 auto-ubiquitination assays and NMR experiments were performed as described previously^{3,15}. Modifications and details are described in Methods. For intrinsic reactivity assays, E2s were charged with Ub for 20–30 min at 37 °C before addition of cysteine, arginine, lysine, serine, threonine or buffer (final concentration 50 mM, pH 7.0). After 15–20 min, reactions were quenched in non-reducing loading-buffer and visualized by Coomassie-stained SDS-PAGE. Reactivity time courses with lysine and cysteine were performed similarly with samples quenched at the indicated times.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 October 2010; accepted 23 February 2011.

Published online 1 May 2011.

1. Anan, T. *et al.* Human ubiquitin-protein ligase Nedd4: expression, subcellular localization and selective interaction with ubiquitin-conjugating enzymes. *Genes Cells* **3**, 751–763 (1998).
2. Zheng, N., Wang, P., Jeffrey, P. D. & Pavletich, N. P. Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases. *Cell* **102**, 533–539 (2000).
3. Brzovic, P. S. *et al.* Binding and recognition in the assembly of an active BRCA1/BARD1 ubiquitin-ligase complex. *Proc. Natl Acad. Sci. USA* **100**, 5646–5651 (2003).
4. Huang, A. *et al.* E2-c-Cbl recognition is necessary but not sufficient for ubiquitination activity. *J. Mol. Biol.* **385**, 507–519 (2009).
5. Capili, A. D., Edghill, E. L., Wu, K. & Borden, K. Structure of the C-terminal RING finger from a RING-IBR-RING/TRIAD motif reveals a novel zinc-binding domain distinct from a RING. *J. Mol. Biol.* **340**, 1117–1129 (2004).
6. Shimura, H. *et al.* Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. *Nature Genet.* **25**, 302–305 (2000).

7. Marin, I., Lucas, J. I., Gradilla, A. C. & Ferrus, A. Parkin and relatives: the RBR family of ubiquitin ligases. *Physiol. Genomics* **17**, 253–263 (2004).
8. Tan, N. G. *et al.* Human homologue of ariadne promotes the ubiquitylation of translation initiation factor 4E homologous protein, 4EHP. *FEBS Lett.* **554**, 501–504 (2003).
9. Tokunaga, F. *et al.* Involvement of linear polyubiquitylation of NEMO in NF- κ B activation. *Nature Cell Biol.* **11**, 123–132 (2009).
10. Whitcomb, E. A., Dudek, E. J., Liu, Q. & Taylor, A. Novel control of S phase of the cell cycle by ubiquitin-conjugating enzyme H7. *Mol. Biol. Cell* **20**, 1–9 (2009).
11. Franssen, K. *et al.* Analysis of SNPs with an effect on gene expression identifies UBE2L3 and BCL3 as potential new risk genes for Crohn's disease. *Hum. Mol. Genet.* **19**, 3482–3488 (2010).
12. Pickart, C. M. & Rose, I. A. Functional heterogeneity of ubiquitin carrier proteins. *J. Biol. Chem.* **260**, 1573–1581 (1985).
13. Wang, X. *et al.* Ubiquitination of serine, threonine, or lysine residues on the cytoplasmic tail can induce ERAD of MHC-1 by viral E3 ligase mK3. *J. Cell Biol.* **177**, 613–624 (2007).
14. Cadwell, K. & Coscoy, L. Ubiquitination of nonlysine residues by a viral E3 ubiquitin ligase. *Science* **309**, 127–130 (2005).
15. Christensen, D. E., Brzovic, P. S. & Kleit, R. E. E2-BRCA1 RING interactions dictate synthesis of mono- or specific polyubiquitin chain linkages. *Nature Struct. Mol. Biol.* **14**, 941–948 (2007).
16. Rodrigo-Brenni, M. C. & Morgan, D. O. Sequential E2s drive polyubiquitin chain assembly on APC targets. *Cell* **130**, 127–139 (2007).
17. Jones, D., Crowe, E., Stevens, T. A. & Candido, E. P. Functional and phylogenetic analysis of the ubiquitylation system in *Caenorhabditis elegans*: ubiquitin-conjugating enzymes, ubiquitin-activating enzymes and ubiquitin-like proteins. *Genome Biology* **3**, RESEARCH0002 (2002).
18. Yunus, A. A. & Lima, C. D. Lysine activation and functional analysis of E2-mediated conjugation in the SUMO pathway. *Nature Struct. Mol. Biol.* **13**, 491–499 (2006).
19. Wu, P. Y. *et al.* A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *EMBO J.* **22**, 5241–5250 (2003).
20. Sakata, E. *et al.* Crystal structure of the UbcH5b~ubiquitin intermediate: insight into the formation of the self-assembled E2~Ub conjugates. *Structure* **18**, 138–147 (2010).
21. Ozkan, E., Yu, H. & Deisenhofer, J. Mechanistic insight into the allosteric activation of a ubiquitin-conjugating enzyme by RING-type ubiquitin ligases. *Proc. Natl Acad. Sci. USA* **102**, 18890–18895 (2005).
22. Kamadurai, H. B. *et al.* Insights into ubiquitin transfer cascades from a structure of UbcH5b~Ubiquitin-HECT^{NEDD4L} complex. *Mol. Cell* **36**, 1095–1102 (2009).
23. Ardley, H. C., Tan, N. G., Rose, S. A., Markhan, A. F. & Robinson, P. A. Features of the parkin/ariadne-like ubiquitin ligase, HHARI, that regulates its interaction with the ubiquitin-conjugating enzyme UbcH7. *J. Biol. Chem.* **276**, 19640–19647 (2001).
24. Eddins, M. J., Carlie, C. M., Gomez, K. M., Pickart, C. M. & Wolberger, C. Mms2-Ubc13 covalently bound to ubiquitin reveals the structural basis of linkage-specific polyubiquitin chain formation. *Nature Struct. Mol. Biol.* **13**, 915–920 (2006).
25. Fallon, L. *et al.* A regulated interaction with the UIM protein Eps15 implicates parkin in EGF receptor trafficking and PI(3)K-Akt signalling. *Nature Cell Biol.* **8**, 834–842 (2006).
26. Joch, M. *et al.* Parkin-mediated monoubiquitination of the PDZ protein PICK1 regulates the activity of acid-sensing ion channels. *Mol. Biol. Cell* **18**, 3105–3118 (2007).
27. Chen, D. *et al.* Parkin mono-ubiquitinates Bcl-2 and regulates autophagy. *J. Biol. Chem.* **285**, 38214–38223 (2010).
28. Maruyama, M. *et al.* Novel mutations, pseudo-dominant inheritance, and possible familial affects in patients with autosomal recessive juvenile parkinsonism. *Ann. Neurol.* **48**, 245–250 (2000).
29. Diao, J., Zhang, Y., Huibregtse, J. M., Zhou, D. & Chen, J. Crystal structure of SopA, a *Salmonella* effector protein mimicking a eukaryotic ubiquitin ligase. *Nature Struct. Mol. Biol.* **15**, 65–70 (2008).
30. Dastur, A., Beaudenon, S., Kellet, M., Krug, R. M. & Huibregtse, M. Herc5, an interferon-induced HECT E3 enzyme, is required for conjugation of ISG15 in human cells. *J. Biol. Chem.* **281**, 4334–4338 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Stoll, M. Schwartz and N. Zheng for critical reading of the manuscript and discussions and V. Vittal, K. Stoll and P. Littlefield for technical help. We thank K. Borden, E. Fon, B. Schulman, L. Boyd, J. Chen and A. Merz for expression vectors for HHARI, parkin, E6-AP, Ubc18, human E1 and GST pull-down reagents, respectively. We acknowledge support from the National Institute of General Medical Sciences in the form of 5R01 GM08055 (R.E.K.) and T32 GM07270 (D.M.W.), and the National Science Foundation in the form of MCB0615632 (R.E.K.)

Author Contributions D.M.W. performed all experiments. P.S.B. and A.L. contributed to the experimental design for Fig. 2 and Supplementary Fig. 5. D.M.W. and R.E.K. designed the overall study and wrote the manuscript with P.S.B.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.E.K. (kleit@u.washington.edu).

METHODS

Multiple sequence alignments. Multiple sequence alignments were performed using Clustal W³¹ with manual sequence adjustments based on E2 structures.

Plasmids, protein expression and purification. Plasmid constructs, protein expression, and purification of wheat E1, Ub, UBC13, UBCH5C, UBE2K, UBCH7, Flag-BRCA1 (residues 1–304)/BARD1 (residues 26–327), Flag-BRCA1 (residues 1–112)/BARD1 (residues 26–140) were described previously^{3,15}. Point mutations were introduced using site-directed mutagenesis (Stratagene) and confirmed by DNA sequencing. PFastBac-His-human E1 was expressed in Hi5 cells and purified by Ni²⁺-affinity chromatography, according to the manufacturer's instructions (Sigma), followed by gel filtration using Superdex 200 resin (GE Health Care Life Sciences). E6-AP (residues 495–852) was expressed and purified as described previously³². Ubc18 was sub-cloned into pHis vector³³ in-frame with the N-terminal His-tag, and UBCH7 was subcloned into pet24a vector in-frame with a His-T7 N-terminal tag. His-T7-UBCH7, His-Ubc18, His-UBCH5C(N77S), HA- and T7-tagged Ub were purified by Ni²⁺ affinity chromatography followed by gel-filtration using Superdex 75 resin (GE Healthcare Life Sciences). Constructs of pGEX-4T parkin (rat) and HHARI were expressed in BL21 *Escherichia coli* (Invitrogen) in Luria Broth supplemented with 0.2 mM ZnCl₂ and were purified using GSTrap FF columns (GE Healthcare and Life Sciences) eluted with 10 mM reduced glutathione. Glutathione was removed by dialysis against 50 mM Tris, 200 mM NaCl, 1 mM DTT, pH 7.6 (parkin) and pH 8.0 (HHARI). HHARI_{R2} was subcloned into pGEX-4T in-frame with the N-terminal GST-tag. The GST-tag on HHARI_{R1-IBR-R2} and HHARI_{R2} was removed by thrombin cleavage for NMR and cysteine modification experiments as well as to repeat activity assays shown in Supplementary Fig. 14.

GST pull-down assays. One-hundred microlitre binding reactions contained GST-HHARI (5 μM) with T7-UBCH7 (5 μM) and 50 μl of glutathione sepharose B resin (GE Healthcare) in the binding buffer: 50 mM Tris, 150 mM NaCl, 0.5% Triton X-100, 0.5 mM dithiothreitol (DTT) pH 7.5. Binding reactions were incubated for 3 h at 4 °C, and resin was washed 5 times with 1.5 ml of binding buffer before proteins were eluted with 80 μl of reduced SDS-PAGE loading buffer. Reaction products were resolved on a 15% SDS-PAGE gel and transferred onto polyvinylidene fluoride membranes (Bio-Rad). The membranes were probed simultaneously with rabbit antibody to GST (Affinity BioReagents) and mouse antibody to T7 (Novagen) followed by goat anti-mouse and goat anti-rabbit secondary antibody conjugated to Alexa Fluor 680 (Molecular probes) and IRdye 800 (Rockland Immunochemicals), respectively. Blotted proteins were detected using an Odyssey infrared imaging system (Licor).

E3 auto-ubiquitination activity assays. One-hundred microlitre reaction mixtures for BRCA1 auto-ubiquitination contained 2 μM His-Flag BRCA1 (residues 1–304)/BARD1 (residues 26–327), 2 μM UBCH5C, 20 μM T7-Ub, 0.5 μM wheat E1 and 10 mM MgCl₂. Reactions were initiated at 37 °C by adding 10 mM ATP and samples were quenched at the indicated time points by boiling in SDS sample buffer that contained βME. Ubiquitination products were visualized by western blot, probing for T7 (Ub) and Flag (BRCA1) epitopes simultaneously. Parkin and HHARI ubiquitination assays were performed similarly except products were visualized by probing for the GST tag on parkin or HHARI or the HA epitope on HA-Ub (mouse primary from Covance). HHARI assays with UBCH7, UBCH5C and UBCH5C(N77S/D87K) mutants were performed at higher concentrations (15 μM E2/E3, 50 μM Ub) and visualized by Coomassie staining on a 15% SDS-PAGE gel. One-hundred microlitre reaction mixtures for E6-AP thioester formation assays included 15 μM E2, 15 μM E6-AP, 30 μM Ub and 10 mM MgCl₂. Reactions were initiated at 37 °C by the addition of 10 mM ATP and gel samples were taken in parallel at the indicated time points in loading buffer that lacked or contained the reducing agent βME.

HHARI thioester detection. Reactions containing 20 μM UBCH7, 20 μM HA-Ub, 0.5 μM E1, 10 mM MgCl₂/ATP were incubated at 37 °C for 30 min to form UBCH7~HA-Ub and chilled on ice. Reactions with HHARI were initiated by diluting 2 μl of charged E2 with 18 μl of 2 μM HHARI_{R1-IBR-R2} on ice and incubating

for the indicated times. Reactions were quenched by the addition of SDS-loading buffer that either contained or lacked the reducing agent βME. HHARI thioesters were visualized by western blot probing for the HA epitope on Ub.

Nucleophile reactivity assays. Reaction mixtures for amino acid reactivity assays contained 20 μM E2, 20 μM Ub, 0.5 μM E1 and 10 mM MgCl₂/ATP in 25 mM NaPi, 150 mM NaCl, pH 7.0 buffer. Amino acids were purchased from Sigma except cysteine (Nutritional Biochemicals Corporation). Five-hundred millimolar stock solutions of L-lysine monohydrochloride, L-arginine monohydrochloride, L-cysteine, L-serine and L-threonine were prepared in reaction buffer and pH was checked by pH paper to be ~7. E2s were charged for 20–30 min at 37 °C before being mixed with 50 mM of cysteine, arginine, lysine, serine, threonine or buffer and incubated for 15–20 min at 37 °C. Samples were quenched in non-reducing loading-buffer and visualized by Coomassie stained SDS-PAGE. Reactivity time-courses with lysine and cysteine were performed similarly except that samples were taken at several time points during the reaction.

In reactions containing E3, 20 μM E3 (His-Flag-BRCA1 residues 1–112 (wild type or I26A)/BARD1 residues 26–140 or E6-AP(C820A) residues 495–852 or GST-HHARI_{R1-IBR-R2}, or GST-HHARI_{R2}) was added and mixed to the pre-charged E2 just before incubation with amino acids. UBCH5C was precharged with Ub where all lysines were mutated to arginines (K0) in reactions with BRCA1 to prevent transfer of Ub. Reactivity reactions visualized by western blot were performed similarly except concentrations were 10 μM E2 or E3, and 5 μM HA-Ub.

HHARI oxyster detection. Ubiquitination reactions contained 15 μM HHARI_{R1-IBR-R2}(C357S), 150 μM HA-Ub, 1.5 μM E1, 10 mM MgCl₂ and 15 μM UBCH7. Reactions were initiated by the addition of 10 mM ATP, and incubated at 37 °C for 30 min before being quenched in reduced SDS-loading buffer. Reactions were then incubated for 20 min at 37 °C with 0.14 N NaOH before being boiled and loaded on a 15% SDS-PAGE gel. Reaction products were visualized by western blot, simultaneously blotting for the HA (Ub) and GST (HHARI) epitopes. For controls, parallel ubiquitination reactions with UBC13 and UBC13(C86S) were performed. UBC13 readily auto-ubiquitinates itself (via an isopeptide) and the UBC13 mutant C86S forms an oxyster-linked Ub conjugate²⁴.

NMR. For the production of ¹⁵N-labelled proteins, bacteria were grown in minimal MOPS medium supplemented with [¹⁵N] ammonium chloride (Cambridge Isotope Labs). NMR data were collected on a Bruker DMX 500 MHz spectrometer. Samples of ¹⁵N-His-Flag-BRCA1(residues 1–112)/BARD1 (residues 26–140) and UBCH5C mutants were prepared as described previously¹⁵. Samples of HHARI_{R2} were prepared as reported previously⁵. Spectra were processed using NMRPipe³⁴/NMRDraw³⁵.

Cysteine modification of HHARI_{R2}. One-hundred microlitre cysteine modification reactions contained 100 μM wild-type HHARI_{R2} and HHARI_{R2}(C357A) and 500 μM 4-(2-iodoacetamido)-TEMPO (Sigma). Stock solutions of 4-(2-iodoacetamido)-TEMPO were prepared at 60 mM in DMSO. Cysteine modification reactions were incubated overnight at 4 °C. Samples for MALDI-TOF were diluted 1:10 in MALDI matrix (saturated sinapinic acid (Sigma) in 40% acetonitrile, 0.1% TFA) and masses were quantified by MALDI-TOF spectrometry on a Bruker AutoFlex II spectrometer, using insulin and apomyoglobin as standards.

- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Huang, L. *et al.* Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. *Science* **286**, 1321–1326 (1999).
- Sheffield, P., Garrard, S. & Derewenda, Z. Overcoming expression and purification problems of RhoGDI using a family of “parallel” expression vectors. *Protein Expr. Purif.* **15**, 34–39 (1999).
- Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
- Johnson, B. A. & Blevins, R. A. NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).

Aberrant lipid metabolism disrupts calcium homeostasis causing liver endoplasmic reticulum stress in obesity

Suneng Fu¹, Ling Yang¹, Ping Li¹, Oliver Hofmann², Lee Dicker², Winston Hide², Xihong Lin², Steven M. Watkins³, Alexander R. Ivanov¹ & Gökhan S. Hotamisligil^{1,4}

The endoplasmic reticulum (ER) is the main site of protein and lipid synthesis, membrane biogenesis, xenobiotic detoxification and cellular calcium storage, and perturbation of ER homeostasis leads to stress and the activation of the unfolded protein response¹. Chronic activation of ER stress has been shown to have an important role in the development of insulin resistance and diabetes in obesity². However, the mechanisms that lead to chronic ER stress in a metabolic context in general, and in obesity in particular, are not understood. Here we comparatively examined the proteomic and lipidomic landscape of hepatic ER purified from lean and obese mice to explore the mechanisms of chronic ER stress in obesity. We found suppression of protein but stimulation of lipid synthesis in the obese ER without significant alterations in chaperone content. Alterations in ER fatty acid and lipid composition result in the inhibition of sarco/endoplasmic reticulum calcium ATPase (SERCA) activity and ER stress. Correcting the obesity-induced alteration of ER phospholipid composition or hepatic *Serca* over-expression *in vivo* both reduced chronic ER stress and improved glucose homeostasis. Hence, we established that abnormal lipid and calcium metabolism are important contributors to hepatic ER stress in obesity.

It has been generally accepted that a surplus of nutrients and energy stimulates synthetic pathways and may lead to client overloading in the ER. However, it has not been demonstrated whether increased *de novo* protein synthesis and client loading into the ER and/or a diminished productivity of the ER in protein degradation or folding leads to ER stress in obesity. Intriguingly, dephosphorylation of eukaryotic translation initiation factor 2 α (eIF2 α) in the liver of high-fat-diet-fed mice reduced the ER stress response³, indicating that additional mechanisms other than translational upregulation may also contribute to ER dysfunction in obesity. To address these mechanistic questions, we first fractionated ER from lean and obese liver tissues (Supplementary Fig. 1a, b) and then extracted ER proteins for comparative proteomic analysis to examine the status of this organelle in obesity. We identified a total of 2,021 unique proteins (Supplementary Table 1). Among them, 120 proteins were differentially regulated in obese hepatic ER samples (Supplementary Fig. 1c and Supplementary Table 2a, b). We independently validated the differential regulation when possible by immunoblot analyses and verified the fidelity of the system (Supplementary Fig. 1d). Gene ontology analysis identified the enrichment of metabolic enzymes—especially ones involved in lipid metabolism—in the obese ER proteome, whereas protein synthesis and transport functions were overrepresented among downregulated ER proteins (Fig. 1a). Consistently, we found that ER-associated protein synthesis was downregulated in the obese liver as demonstrated by polysome profiling (data not shown), whereas the expression of genes involved in *de novo* lipogenesis (*Fas*, *Scd1*, *Ces1d*, *Dgat2* and *Dak*) and phospholipid synthesis (*Pcyt1a* and *Pemt*) were broadly upregulated (Fig. 1b, c).

We also observed upregulation of protein degradation pathways but did not find a broad change in the quantity of ER chaperones (Supplementary Fig. 2 and Supplementary Table 2a). Taken together, these data revealed a fundamental shift in hepatic ER function in obesity from protein to lipid synthesis and metabolism.

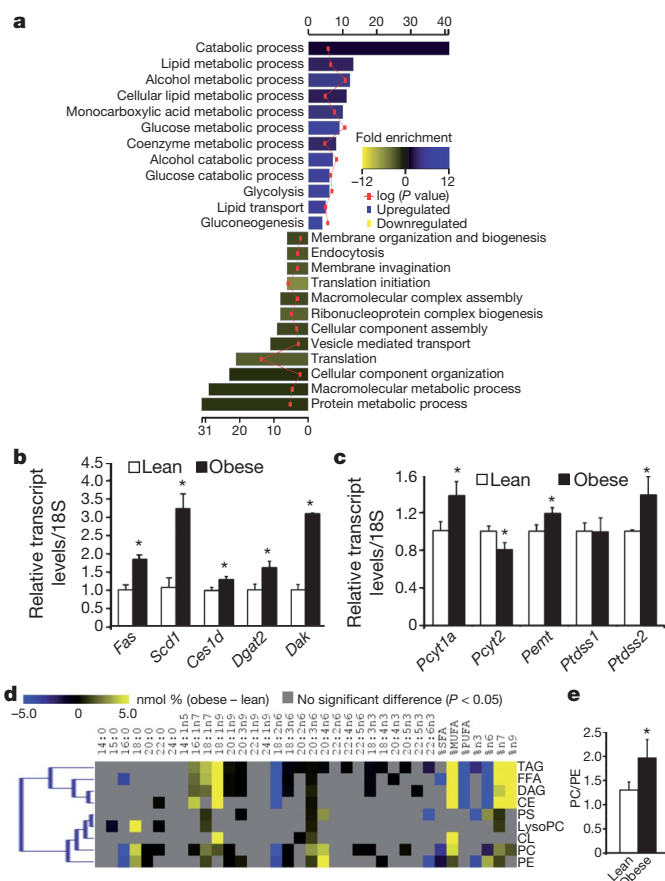


Figure 1 | Proteomic and lipidomic landscape of the lean and obese ER. **a**, Biological pathways associated with significantly regulated proteins in the obese ER proteome. Bar colours indicate the fold enrichment with significance values (negative log of P values) superimposed. **b**, **c**, Transcript levels of genes involved in lipid metabolism in the lean and obese mouse liver. **d**, Alterations of liver ER lipidome. Heatmap display of all significant ($P < 0.05$) alterations present between lean and obese ER lipidomes. The colour corresponds to differences in the relative abundance (nmol percentage) of each fatty acid among individual lipid groups detected in the lean and obese liver ER. **e**, The relative abundance of PC and PE in lean and obese liver ER samples. Values are mean \pm s.e.m. ($n = 6$ for each group). * $P < 0.05$, Student's t -test.

¹Departments of Genetics and Complex Diseases, and Nutrition, Harvard School of Public Health, Boston, Massachusetts 02115, USA. ²Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA. ³Lipomics Technologies Inc, West Sacramento, California 95691, USA. ⁴Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA.

The presence of chronic ER stress in obese liver (Supplementary Fig. 2) despite a reduction in ER-associated protein synthesis led us to postulate that the ER stress in obesity may not simply be invoked by protein overloading but also driven by compromised folding capacity, in which lipid metabolism may have a function⁴. For example, the ability of palmitate and cholesterol to induce ER stress in cultured cells correlates with their incorporation into the ER^{5,6}. Therefore, we quantitatively determined all major lipid species and their fatty acid composition in ER samples isolated from lean and obese liver along with the diet consumed by these mice (Supplementary Fig. 3 and Supplementary Table 3). First, we found that the fatty acid composition of ER lipids in the lean mouse liver was distinct from corresponding dietary lipids, indicating the contribution of a basal level of *de novo* lipogenesis to the biogenesis of ER membranes *in vivo* (Supplementary Fig. 3a, b and Supplementary Table 3). Almost all ER-derived lipids were composed of significantly higher levels of saturated fatty acids (SFAs) whereas their polyunsaturated fatty acid (PUFA) content was much lower than those of corresponding dietary lipids, indicating that *de novo* synthesized SFAs are preferred over diet-derived PUFAs as the substrate for the synthesis of hepatic ER lipids. Second, the liver ER samples of lean and obese mice also had profoundly different compositions of fatty acids and lipids as illustrated by the clear separation of lean and obese ER lipidome in cluster analysis (Supplementary Fig. 3c). The obese ER was significantly enriched with monounsaturated fatty acids (MUFAs; Fig. 1d), a bona fide product of *de novo* lipogenesis, in liver. Third, the obese ER samples contained a higher level of phosphatidylcholine (PC) as compared to phosphatidylethanolamine (PE) (PC/PE = 1.97 versus 1.3, $P < 0.05$; Fig. 1e and Supplementary Table 3), two of the most abundant phospholipids on the ER membrane. The rise of the PC/PE ratio is probably caused by the upregulation of two key genes involved in PC synthesis and PE to PC conversion: choline-phosphate cytidylyltransferase A (*Pcyt1a*) and phosphatidylethanolamine *N*-methyltransferase (*Pemt*) (Fig. 1c and Supplementary Fig. 3a), and it is consistent with the essential role of PC for lipid packaging in the form of lipid droplets or lipoproteins, both of which are increased in obesity. In contrast, the PC/PE ratio in the lean hepatic ER was essentially identical as it is in the diet (Supplementary Table 3), indicating that the increase of PC/PE ratio in obesity is not due to food consumption, but the result of increased lipid synthesis in the obese liver.

The desaturation of SFAs to MUFAs in the obese liver probably has a protective role in reducing lipotoxicity, whereas the decrease of PUFA content in the ER may limit its reducing capacity and contribute to ER stress⁷. However, a potential role of the PC/PE ratio in regulating ER homeostasis has not been studied before. Previous biochemical studies have shown that increasing PC content in the membrane inhibits the calcium transport activity of SERCA^{5,8}. Consistently, we found that the addition of PC to liver-derived microsomes *in vitro* substantially inhibited SERCA activity (Fig. 2a). More importantly, overexpression of the PE to PC conversion enzyme *Pemt* in Hepa1-6 cells significantly inhibited microsomal SERCA activity, indicating that changes in the PC/PE balance in a cellular setting can significantly perturb SERCA function (Fig. 2b, c). As calcium has an important role in mediating chaperone function and protein folding in the ER, and given that SERCA is principally responsible for maintaining calcium homeostasis in this organelle, we postulated that the increased PC/PE ratio in the ER of obese liver might impair ER calcium retention and homeostasis *in vivo*, thereby contributing to protein misfolding and ER stress. In support of this possibility, we found that the calcium transport activity of microsomes prepared from the livers of obese mice was significantly lower than those isolated from lean animals (4.6 ± 0.2 versus 5.3 ± 0.3 , $P = 0.046$; Fig. 2d), despite the fact that the SERCA protein level was modestly higher in the former, consistent with an inhibitory role of the PC/PE ratio on SERCA function.

Modest defects in SERCA activity have been implicated in the pathology of Darier's disease⁹, and we found that a reduction in SERCA expression *in vivo* (Fig. 2e) and a concurrent reduction in its calcium

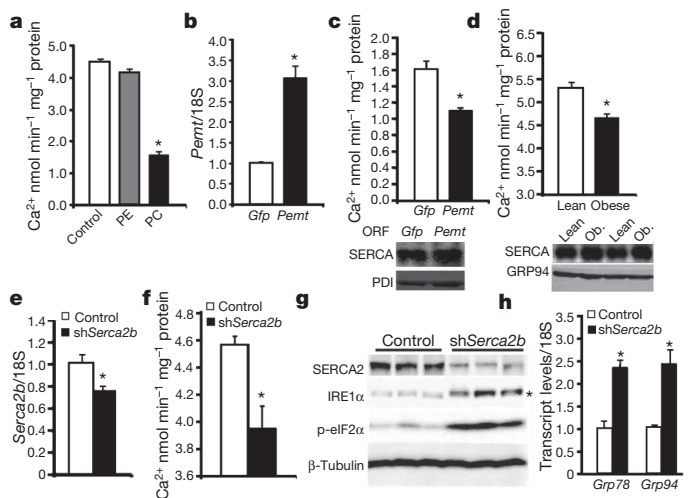


Figure 2 | An increased PC/PE ratio impairs SERCA activity and ER homeostasis. **a**, Calcium transport activity of microsomes loaded with PC and PE *in vitro*. **b**, **c**, Transcript levels of *Pemt* (**b**) and corresponding microsomal calcium transport activities (**c**) of Hepa1-6 cells expressing control (*Gfp*) or mouse *Pemt* open reading frames (ORFs). **d**, Calcium transport activity (top) and SERCA protein levels (bottom) of microsomes prepared from lean and obese mouse liver. **e–h**, Liver *Serca2b* transcript levels (**e**) and microsomal calcium transport activities (**f**), immunoblot (**g**) and quantitative RT–PCR (**h**) measurement of ER stress markers in the livers of lean mice expressing either *LacZ* (control) or *Serca2b* shRNAs. Asterisk in **g** denotes the phosphorylated IRE1α and in other panels denotes significant difference ($*P < 0.05$, $n = 4$) by Student's *t*-test. Values are mean \pm s.e.m.

transport activity (Fig. 2f) potentially activated hepatic ER stress in lean mice as evidenced by IRE1α and eIF2α phosphorylation and changes in the expression of GRP78 and GRP94 (Fig. 2g, h). Therefore, there seems to be little redundancy in the function of SERCA beyond physiological fluctuations to maintain ER homeostasis, and the reduction in calcium transport activity could be a potential mechanism of hepatic ER stress in obesity.

We carried out two different but complementary approaches to correct aberrant lipid metabolism induced SERCA dysfunction and examined the effects on ER homeostasis in the obese liver. If the alteration in PC/PE ratio seen in obese liver is a significant contributor to ER stress, correction of this ratio to lean levels by reducing *Pemt* expression should improve calcium transport defects and produce beneficial effects on hepatic ER stress and metabolism. Using an adenovirally expressed short hairpin RNA (shRNA), we were able to achieve ~50–70% suppression of the *Pemt* transcript in obese liver (Supplementary Fig. 4a). As postulated, suppression of *Pemt* led to a decrease of PC content from ~39% to ~33%, which was compensated by an ~7% increase of PE content from ~17% to 24% (Supplementary Table 4). As a result, the PC/PE ratio is reduced to 1.3 (equivalent to the lean ratio), as compared to 2.0 detected in the ER of the obese liver (Fig. 3a). The reduction of the PC/PE ratio was accompanied by a significant improvement in the calcium transport activity of the ER prepared from the *Pemt*-knockdown obese mice (Fig. 3b). As the improvement of calcium transport function occurred with few and minor changes in the overall fatty acid composition of ER (Supplementary Fig. 4b, c and Supplementary Table 5), our results confirmed the rise in PC/PE ratio as an inhibitory factor of SERCA activity in obesity. More importantly, hepatic ER stress indicators including the phosphorylation of IRE1α and eIF2α as well as the expression of C/EBP homologous protein (CHOP), homocysteine-inducible, ER stress-inducible protein (HERP) and Der1-like domain family member 2 (DERL2) were all reduced upon suppression of *Pemt* in obese mice (Fig. 3c, d and Supplementary Fig. 4d). Relief of chronic ER stress in leptin-deficient (*Lep*^{-/-}) mice has been associated with improvement of hepatic steatosis and glucose homeostasis^{10,11}.

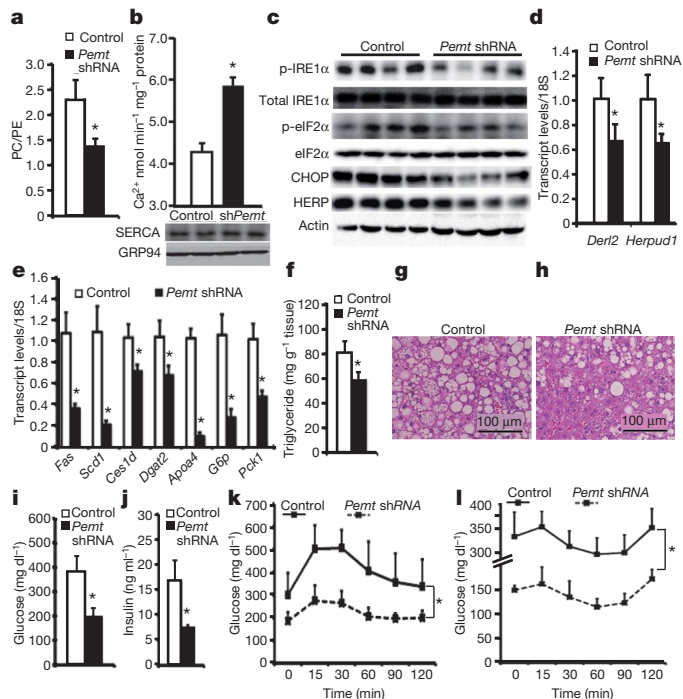


Figure 3 | Suppression of liver *Pemt* expression corrects the ER PC/PE ratio, relieves ER stress and improves systemic glucose homeostasis in obesity. a, b, PC/PE ratio (a) and calcium transport activity (b) of liver ER from *Lep^{-/-}* mice expressing *LacZ* (control) or *Pemt* shRNAs. c, d, Immunoblot (c) and quantitative PCR (d) measurement of ER stress markers in the liver. e–h, Expression of hepatic lipogenesis and gluconeogenesis genes (e), triglyceride content (f) and haematoxylin & eosin staining (g and h) of liver samples. i, j, Plasma glucose (i) and insulin (j) levels in control and *Pemt* shRNA-treated *Lep^{-/-}* mice after 6 h food withdrawal. k, l, Plasma glucose levels of control and *Pemt* shRNA-treated *Lep^{-/-}* mice after intraperitoneal administration of either 1 g kg⁻¹ of glucose (k) or 1 IU kg⁻¹ of insulin (l). All data are mean ± s.e.m. (*n* = 4 for a–e, *n* = 6 for f–l). **P* < 0.05 (one-way ANOVA for data presented in k and l, and Student's *t*-test for others).

Consistently, genes involved in hepatic lipogenesis (*Fas*, *Scd1*, *Ces1d*, *Dgat2*) and lipoprotein synthesis (*Apoa4*) were significantly downregulated in the obese liver after suppression of *Pemt* (Fig. 3e). As a result, these mice exhibited a significant reduction in hepatic steatosis and liver triglyceride content (Fig. 3f–h). Genes involved in glucose production (*G6pc*, *Pck1*) in the liver were significantly downregulated (Fig. 3e), and there were also significant reductions in both hyperglycaemia and hyperinsulinaemia in obese mice after the suppression of hepatic *Pemt* expression (Fig. 3i, j). Glucose and insulin tolerance tests revealed significantly enhanced glucose disposal after *Pemt* suppression (Fig. 3k, l). A similar phenotype is also observed upon suppression of hepatic *Pemt* in high-fat-diet-induced obesity, with reduced ER stress and improved glucose homeostasis (Supplementary Fig. 5). These data are consistent with the phenotype seen in *Pemt*-deficient mice, which exhibit protection against diet-induced insulin resistance and atherosclerosis¹². Therefore, correcting the PC/PE ratio of the ER can significantly improve calcium transport defects, reduce ER stress and improve metabolism, supporting the hypothesis that changes in lipid metabolism contribute to SERCA dysfunction, ER stress and hyperglycaemia in both genetic- and diet-induced models of obesity.

We then carried out overexpression of hepatic *Serca* *in vivo* to overcome the partial inhibition of SERCA activity by PC (Fig. 4a). Indeed, exogenous SERCA expression in the liver of *Lep^{-/-}* mice improved the calcium import activity of the ER (Fig. 4b), restored euglycaemia and normoinsulinaemia within a few days, and markedly improved glucose tolerance (Fig. 4c, d and Supplementary Fig. 6). Upon *Serca* expression, the liver showed an increase in size but a marked reduction of lipid infiltration (Fig. 4e–h) and suppression of

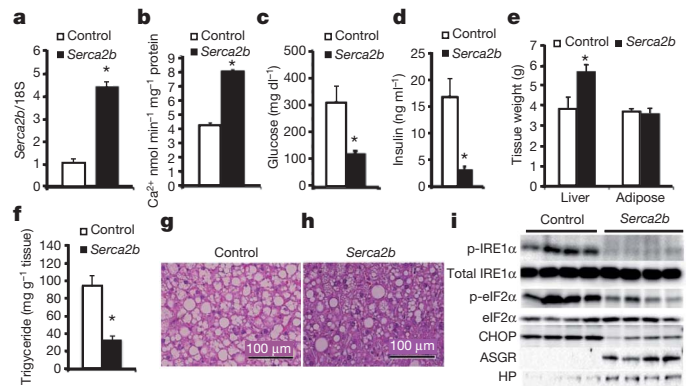


Figure 4 | Exogenous *Serca* expression alleviates ER stress and improves systemic glucose homeostasis. a, b, Liver *Serca2b* transcript levels (a) and microsomal calcium transport activities (b) of control or *Serca2b*-overexpressing obese mice. c–e, Plasma glucose (c), plasma insulin levels (d) and tissue weights (e) of *Lep^{-/-}* mice as in panel a. f–i, Triglyceride content (f), haematoxylin & eosin staining (g, h) and immunoblot analyses (i) of ER stress markers (IRE1α and eIF2α phosphorylation, and CHOP) and secretory proteins (ASGR and HP) in the obese liver expressing *Serca2b* compared to controls. All values are mean ± s.e.m. (*n* = 4 for a and b, *n* = 6 for c–h). **P* < 0.05 (Student's *t*-test).

IRE1α and eIF2α phosphorylation, along with a significant reduction in CHOP levels (Fig. 4i). In these liver samples, there was also a marked increase in two secretory proteins that were otherwise diminished in obesity: asialoglycoprotein receptor (ASGR) and haptoglobin (HP) (Fig. 4i). As the folding and maturation of ASGR is most sensitive to perturbations of calcium homeostasis in the ER¹³, our results indicate that exogenously increased SERCA expression restored calcium homeostasis and relieved at least some aspects of chronic ER stress in the obese liver. Taken together, these data reinforce the hypothesis that lipid-driven alterations and ER calcium homeostasis are important contributors to hepatic ER stress in obesity.

The chronic activation of ER stress markers has been observed in a variety of experimental obese models as well as in obese humans¹⁴. Furthermore, treatment of obese mice and humans with chemical chaperones results in increased insulin sensitivity^{10,15}. Our systematic, compositional and functional characterization of hepatic ER landscape from lean and obese mice revealed a diametrically opposite regulation of ER functions regarding protein and lipid metabolism and revealed mechanisms giving rise to ER stress. In particular, an increase in the PC/PE ratio in the ER, driven by the upregulation of *de novo* lipogenesis in obesity, was linked to SERCA dysfunction and chronic ER stress *in vivo*. During the review of this manuscript, a study reported downregulation of the SERCA protein level in obese liver¹⁶, which was not evident in our analysis and seemed to have resulted from the choice of methodology in ER protein preparations (Supplementary Fig. 7). Nevertheless, other mechanisms such as oxidative and inflammatory changes associated with obesity can also perturb ER homeostasis by affecting ER calcium fluxes^{17–19} and will be important to study in the future.

The identification of a lipid-driven calcium transport dysfunction and ER stress provides a fundamental framework for understanding the pathogenesis of hepatic lipid metabolism and chronic ER stress in obesity. First, excessive food intake inevitably stimulates lipogenesis for energy storage, and PC is the preferred phospholipid coat of lipid droplets and lipoproteins²⁰. Therefore, there is a biological need for the synthesis of more PC for packaging and storing the products of hepatic lipogenesis. Second, *de novo* fatty acid synthesis in the obese liver produces ample amounts of MUFA, which is effectively incorporated into PC but not PE, which further distorts the PC/PE ratio and impairs ER function. The resulting ER stress facilitates the secretion of excessive lipids from the liver without ameliorating hyperinsulinaemia-induced lipogenesis²¹, and thus hepatosteatosis and ER stress ensue. As a result,

relieving ER stress in obesity may ultimately depend on breaking this 'lipogenesis-ER-stress-lipogenesis' vicious cycle and restoring ER folding capacity. Therefore, we suggest that genetic, chemical or dietary interventions that modulate hepatic phospholipid synthesis and/or ER calcium homeostasis function might represent a new set of therapeutic opportunities for common chronic diseases associated with ER stress, such as obesity, insulin resistance and type 2 diabetes.

METHODS SUMMARY

Male leptin-deficient (*Lep*^{-/-}) and wild-type littermates in the C57BL/6J background were either bred in-house or purchased from the Jackson Laboratory (strain B6.V-Lep^{ob/J}, stock number 000632). Transduction of adenoviruses (serotype 5, Ad5) for the expression of open reading frames (ORFs) or shRNAs was carried out between 10–11 weeks after birth, and all mice were killed between 12–13 weeks of age, unless noted otherwise. ER fractionation for proteomic and lipidomic analysis were carried out as previously described²². Calcium transport experiments were performed as previously described²³, with some modifications. Quantitative RT-PCR, western blot analysis, histology and *in vivo* animal experiments were carried out as previously described^{10,24}. Oligonucleotide sequences used in this study are listed in Supplementary Table 6. Detailed experimental procedures and protocols are described in the Supplementary Material.

Received 13 October 2010; accepted 22 February 2011.

Published online 1 May 2011.

- Ron, D. & Walter, P. Signal integration in the endoplasmic reticulum unfolded protein response. *Nature Rev. Mol. Cell Biol.* **8**, 519–529 (2007).
- Hotamisligil, G. S. Endoplasmic reticulum stress and the inflammatory basis of metabolic disease. *Cell* **140**, 900–917 (2010).
- Oyadomari, S. *et al.* Dephosphorylation of translation initiation factor 2 α enhances glucose tolerance and attenuates hepatosteatosis in mice. *Cell Metab.* **7**, 520–532 (2008).
- Erbay, E. *et al.* Reducing endoplasmic reticulum stress through a macrophage lipid chaperone alleviates atherosclerosis. *Nature Med.* **15**, 1383–1391 (2009).
- Li, Y. *et al.* Enrichment of endoplasmic reticulum with cholesterol inhibits sarcoplasmic-endoplasmic reticulum calcium ATPase-2b activity in parallel with increased order of membrane lipids: implications for depletion of endoplasmic reticulum calcium stores and apoptosis in cholesterol-loaded macrophages. *J. Biol. Chem.* **279**, 37030–37039 (2004).
- Borradaile, N. M. *et al.* Disruption of endoplasmic reticulum structure and integrity in lipotoxic cell death. *J. Lipid Res.* **47**, 2726–2737 (2006).
- Kim, S. J. *et al.* Omega-3 and omega-6 fatty acids suppress ER- and oxidative stress in cultured neurons and neuronal progenitor cells from mice lacking PPT1. *Neurosci. Lett.* **479**, 292–296 (2010).
- Cheng, K. H., Lepock, J. R., Hui, S. W. & Yeagle, P. L. The role of cholesterol in the activity of reconstituted Ca-ATPase vesicles containing unsaturated phosphatidylethanolamine. *J. Biol. Chem.* **261**, 5081–5087 (1986).
- Miyauchi, Y. *et al.* Comprehensive analysis of expression and function of 51 sarco(endo)plasmic reticulum Ca²⁺-ATPase mutants associated with Darier disease. *J. Biol. Chem.* **281**, 22882–22895 (2006).
- Ozcan, U. *et al.* Chemical chaperones reduce ER stress and restore glucose homeostasis in a mouse model of type 2 diabetes. *Science* **313**, 1137–1140 (2006).
- Kammoun, H. L. *et al.* GRP78 expression inhibits insulin and ER stress-induced SREBP-1c activation and reduces hepatic steatosis in mice. *J. Clin. Invest.* **119**, 1201–1215 (2009).
- Jacobs, R. L. *et al.* Impaired *de novo* choline synthesis explains why phosphatidylethanolamine N-methyltransferase-deficient mice are protected from diet-induced obesity. *J. Biol. Chem.* **285**, 22403–22413 (2010).
- Lodish, H. F. & Kong, N. Perturbation of cellular calcium blocks exit of secretory proteins from the rough endoplasmic reticulum. *J. Biol. Chem.* **265**, 10893–10899 (1990).
- Gregor, M. F. *et al.* Endoplasmic reticulum stress is reduced in tissues of obese subjects after weight loss. *Diabetes* **58**, 693–700 (2009).
- Kars, M. *et al.* Tauroursodeoxycholic acid may improve liver and muscle but not adipose tissue insulin sensitivity in obese men and women. *Diabetes* **59**, 1899–1905 (2010).
- Park, S. W., Zhou, Y., Lee, J. & Ozcan, U. Sarco(endo)plasmic reticulum Ca²⁺-ATPase 2b is a major regulator of endoplasmic reticulum stress and glucose homeostasis in obesity. *Proc. Natl Acad. Sci. USA* **107**, 19320–19325 (2010).
- Li, S. Y. *et al.* Cardiac contractile dysfunction in Lep/Lep obesity is accompanied by NADPH oxidase activation, oxidative modification of sarco(endo)plasmic reticulum Ca²⁺-ATPase and myosin heavy chain isozyme switch. *Diabetologia* **49**, 1434–1446 (2006).
- Cardozo, A. K. *et al.* Cytokines downregulate the sarcoendoplasmic reticulum pump Ca²⁺ ATPase 2b and deplete endoplasmic reticulum Ca²⁺, leading to induction of endoplasmic reticulum stress in pancreatic β -cells. *Diabetes* **54**, 452–461 (2005).
- Li, G. *et al.* Role of ER01- α -mediated stimulation of inositol 1,4,5-triphosphate receptor activity in endoplasmic reticulum stress-induced apoptosis. *J. Cell Biol.* **186**, 783–792 (2009).
- Schiller, J. *et al.* Lipid analysis of human HDL and LDL by MALDI-TOF mass spectrometry and (31)P-NMR. *J. Lipid Res.* **42**, 1501–1508 (2001).
- Brown, M. S. & Goldstein, J. L. Selective versus total insulin resistance: a pathogenic paradox. *Cell Metab.* **7**, 95–96 (2008).
- Cox, B. & Emili, A. Tissue subcellular fractionation and protein extraction for use in mass-spectrometry-based proteomics. *Nature Protocols* **1**, 1872–1878 (2006).
- Moore, L., Chen, T., Knapp, H. R. Jr & Landon, E. J. Energy-dependent calcium sequestration activity in rat liver microsomes. *J. Biol. Chem.* **250**, 4562–4568 (1975).
- Cao, H. *et al.* Identification of a lipokine, a lipid hormone linking adipose tissue to systemic metabolism. *Cell* **134**, 933–944 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Porter, E. Freeman and R. Davis for technical assistance. The anti-HERP antibody is a gift of Y. Hirabayashi. We thank the members of the G.S.H. laboratory for scientific discussions and critical reading of the manuscript. This work was supported in part by the National Institutes of Health (DK52539 and 1RC4-DK090942) and a research grant from Syndexa Pharmaceuticals to G.S.H. S.F. was supported in part by the NIH/NIEHS postdoctoral training grant (T32ES007155).

Author Contributions S.F. designed, performed experiments, analysed and interpreted the results and wrote the manuscript; L.Y. and P.L. performed some animal experiments; O.H., L.D., W.H. and X.L. performed statistical and bioinformatic analysis of the proteomic data; S.W.M. quantified the lipid composition of ER and analysed the data; A.R.I. analysed the protein composition of ER; G.S.H. generated the hypothesis, designed the project, analysed and interpreted the data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.S.H. (ghotamis@hsph.harvard.edu).

Sequential interactions with Sec23 control the direction of vesicle traffic

Christopher Lord^{1*}, Deepali Bhandari^{1*}, Shekar Menon¹, Majid Ghassemian², Deborah Nycz³, Jesse Hay³, Pradipta Ghosh⁴ & Susan Ferro-Novick¹

How the directionality of vesicle traffic is achieved remains an important unanswered question in cell biology. The Sec23p/Sec24p coat complex sorts the fusion machinery (SNAREs) into vesicles as they bud from the endoplasmic reticulum (ER). Vesicle tethering to the Golgi begins when the tethering factor TRAPPI binds to Sec23p. Where the coat is released and how this event relates to membrane fusion is unknown. Here we use a yeast transport assay to demonstrate that an ER-derived vesicle retains its coat until it reaches the Golgi. A Golgi-associated kinase, Hrr25p (CK1 δ orthologue), then phosphorylates the Sec23p/Sec24p complex. Coat phosphorylation and dephosphorylation are needed for vesicle fusion and budding, respectively. Additionally, we show that Sec23p interacts in a sequential manner with different binding partners, including TRAPPI and Hrr25p, to ensure the directionality of ER–Golgi traffic and prevent the back-fusion of a COPII vesicle with the ER. These events are conserved in mammalian cells.

Membrane fusion is mediated by a highly conserved family of membrane proteins called SNAREs. The pairing of a SNARE on the vesicle (v-SNARE) with its cognate SNARE on the target membrane (t-SNARE) is required for fusion¹; however, the same v-SNARE (Sec22p) can act in both anterograde ER–Golgi and retrograde Golgi–ER traffic². This observation implies that factors other than the SNAREs define the direction of membrane flow.

Motifs in the ER–Golgi SNAREs needed for fusion are masked by Sec24p (subunit of the COPII coat) as these fusogens are sorted into ER-derived vesicles³. The COPII coat is assembled on the ER when the activated form of the GTPase Sar1p (Sar1p–GTP) recruits the Sec23p/Sec24p complex by binding to Sec23p, the GTPase-activating protein (GAP) for Sar1p⁴. Polymerization of the coat requires the recruitment of the Sec13p/Sec31p complex (coat outer shell) by the Sec23p/Sec24p complex, which leads to the hydrolysis of GTP on Sar1p and vesicle fission⁴.

The initial interaction of a vesicle with its target membrane is mediated by a class of proteins called tethers that work in conjunction with GTPases of the Rab family⁵. The tethering factor TRAPPI is a multimeric guanine nucleotide exchange factor (GEF) that recruits and activates the Rab GTPase Ypt1p⁶. Previous findings showed that the interaction of TRAPPI with the coat adaptor protein Sec23p is required for vesicle tethering⁶. These studies, however, did not address the question of whether the coat is shed before or after the vesicles bind to the Golgi. Here we show that a Golgi-localized kinase, Hrr25p, displaces purified TRAPPI that is pre-bound to Sec23p and phosphorylates the Sec23p/Sec24p complex. Our findings show that the COPII coat subunit Sec23p interacts in a hierarchical manner with Sar1p, TRAPPI and Hrr25p to ensure the directionality of anterograde membrane flow.

The Golgi inhibits TRAPPI vesicle binding

After COPII vesicles bud from the ER, Sar1p is released from vesicles when GTP is hydrolysed, but the inner and outer shells of the coat are largely retained (Supplementary Fig. 1)⁷. To define the events that

occur after TRAPPI binds to Sec23p, we immobilized TRAPPI on beads and asked when COPII vesicles lose their ability to bind. For these studies, the binding of pro- α -factor-containing vesicles formed with cytosol was considered to be 100% (Fig. 1a). We observed that vesicles formed in the presence of Golgi lost their ability to bind TRAPPI (Fig. 1a). Because the binding of vesicles to TRAPPI is mediated by the COPII coat, this experiment indicates that COPII vesicles lose their ability to bind TRAPPI because the Golgi contains a factor that either releases or modifies the coat. To determine whether vesicles must tether to the Golgi to lose their ability to bind to TRAPPI, we formed vesicles with *bet3-1* mutant fractions. The *bet3-1* mutant, which harbours a mutation in the Bet3p subunit of TRAPPI, is defective in vesicle tethering⁶. The defect in this mutant is partially complemented *in vitro* by the addition of purified recombinant TRAPPI (Supplementary Fig. 2a). Vesicles formed from *bet3-1* donor cells and cytosol, with or without Golgi, bound equally well to the TRAPPI-containing beads (Fig. 1a). These findings indicate that the vesicles must tether to the Golgi to lose their ability to bind to TRAPPI.

COPII vesicle tethering requires TRAPPI, Ypt1p and Uso1p^{5,6}. To determine when COPII vesicles lose their ability to bind to TRAPPI, we blocked vesicle tethering and fusion at several different steps *in vitro* in the presence of Golgi membranes. The pro- α -factor-containing membranes, formed during these blocks, were then tested for their ability to bind to TRAPPI. The transport incompetent vesicles that formed, when Ypt1p function was blocked with anti-Ypt1p antibody, bound efficiently to TRAPPI (Fig. 1b). Disrupting Ypt1p function should also block the recruitment to vesicles of the long coiled-coil tether Uso1p (yeast orthologue of p115), a Ypt1p effector that links donor and acceptor membranes to each other *in vitro*⁵. When we formed vesicles with fractions from the *uso1-1* mutant, transport incompetent vesicles retained their ability to bind TRAPPI at 27 °C and 17 °C (Fig. 1c).

We also blocked fusion *in vitro* with antibody directed against Sec22p (SNARE) or Sly1p, a Sec1-like protein that binds to SNAREs⁸. *In vitro*, neutralizing antibody to Sly1p blocks *trans*-SNARE complex formation⁸. Binding to TRAPPI decreased when vesicles were formed in the

¹Department of Cellular and Molecular Medicine, Howard Hughes Medical Institute, University of California at San Diego, La Jolla, California 92093-0668, USA. ²Department of Chemistry and Biochemistry, Biomolecular and Proteomics Mass Spectrometry Facility, University of California at San Diego, La Jolla, California 92093, USA. ³Division of Biological Sciences, The University of Montana, Missoula, Montana 59812, USA. ⁴Department of Medicine, University of California at San Diego, La Jolla, California, 92093-0651, USA.

*These authors contributed equally to this work.

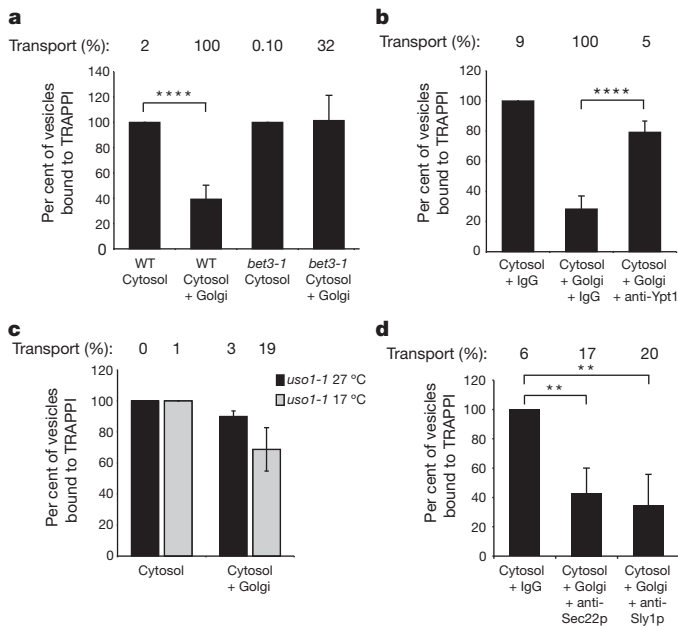


Figure 1 | COPII vesicles lose their ability to bind TRAPPI after Uso1p acts. **a**, Vesicles formed *in vitro* with cytosol with or without Golgi were incubated with TRAPPI-containing beads and the precipitated radiolabelled cargo was counted. Cytosol and Golgi were isolated from either wild type (WT) or the *bet3-1* mutant. Error bars represent standard deviation (s.d.), $N = 6$. In **a–d**, the per cent transport observed for the total reaction is reported above the bar graphs. **b**, **d**, Cytosol and Golgi fractions, derived from wild type, were formed with: anti-Ypt1p (12 μ g), anti-Sec22p (12 μ g), anti-Sly1p (20 μ g) antibodies, or IgG (12 μ g or 20 μ g, respectively). Error bars represent s.d., $N = 4$. **c**, Vesicles were formed with *uso1-1* fractions at 27 °C or 17 °C and incubated with TRAPPI-containing beads. $N = 2$, bars show the range. ** $P < 0.01$, **** $P < 0.0001$ Student's *t*-test.

presence of anti-Sec22p or anti-Sly1p antibodies (Fig. 1d). These findings indicate that COPII vesicles lose their ability to bind TRAPPI after Uso1p functions, but before *trans*-SNARE complex formation. Uso1p does not seem to have a role in vesicle uncoating, as the membrane and soluble pools of Sec23p were unaltered in the *uso1-1* mutant *in vivo* (Supplementary Fig. 3a) and Uso1p/p115 did not release Sec23 from membranes *in vitro* (see Supplementary Figs 3b, c). Together, these findings indicate that COPII vesicles retain their coat until they reach the Golgi.

Hrr25p phosphoregulates Sec23p/Sec24p

Despite the fact that only $36 \pm 0.65\%$ of wild-type vesicles uncoat *in vitro* in the presence of Golgi membranes (Supplementary Fig. 3d), $61 \pm 4\%$ of the vesicles lose their ability to bind to TRAPPI (Fig. 1a). This observation indicates that the inability of COPII vesicles to bind TRAPPI is not just a consequence of vesicle uncoating. Because the COPII coat inner shell is known to be phosphorylated in mammalian cells⁹, we considered the possibility that phosphorylation of Sec23p may block the ability of COPII vesicles to bind TRAPPI.

To identify a kinase that could phosphorylate Sec23p, we searched the yeast database (<http://www.yeastgenome.org>) for an essential (Supplementary Fig. 3e) Golgi-localized kinase. Only one kinase, Hrr25p, was found to have an orthologue (CKI δ , human orthologue of casein kinase I δ) that localizes at the Golgi and ER–Golgi interface in mammalian interphase cells¹⁰. Inhibiting CKI δ function was reported to block ER–Golgi traffic¹¹, and a mutation that reduced the kinase activity of Hrr25p was shown to suppress the temperature-sensitive COPII vesicle budding defect in the *sec12-4* mutant¹². Although these results implicated Hrr25p/CKI δ in ER–Golgi traffic, its role in membrane traffic is not well defined.

Hrr25p is known to reside in the nucleus in G1-arrested cells and, like CKI δ , it is found at the spindle pole body (SPB) in nocodazole-treated (M-phase) cells^{10,13}. When we visualized Hrr25p–RFP (genomic copy tagged with red fluorescent protein) in asynchronously grown cells, however, the majority (>95%) was found on punctate structures that largely colocalize with early (Vrg4p) and late Golgi (Sec7p) markers (Fig. 2a). Occasionally, we observed Hrr25p–RFP in the nucleus, at puncta along the nuclear envelope (presumably the SPB), and at the bud neck and cortex, as previously reported^{13,14}. Consistent with its Golgi localization, all of the haemagglutinin (HA)-tagged Hrr25p co-fractionated with membranes in differential fractionation experiments (Fig. 5a, bottom).

Both glutathione S-transferase (GST)-tagged Sec23p and GST–Sec24p were phosphorylated by His₆-tagged Hrr25p (Fig. 2b and Supplementary Fig. 3f, compare lanes 1 and 2), but not catalytically inactive His₆–Hrr25p(K38A) *in vitro* (Fig. 2b and Supplementary Fig. 3f, lane 3). Phosphorylation *in vivo* was examined with a conditional allele of *hrr25* (ref. 14). In this mutant, HA-tagged HRR25^{degron} was placed behind the inducible GAL promoter as the sole copy of the gene. When this strain is grown in galactose, HA–Hrr25p^{degron} is expressed. In the presence of glucose, however, the expression of HA–Hrr25p^{degron} ceases and the protein is rapidly degraded. After 10 h in glucose, when a delay in trafficking of carboxypeptidase Y between the ER–Golgi was observed, most of the Hrr25p^{degron} was degraded (not shown). Lysates prepared from the *hrr25* mutant grown with galactose (+Hrr25p) or glucose (–Hrr25p) (Supplementary Fig. 3g) were precipitated with anti-Sec24p (Fig. 2c, lanes 1 and 3), or IgG (Fig. 2c, lanes 2 and 4). Western blot analysis of the immunoprecipitates with anti-phospho-Ser/Thr antibody revealed that phospho-Sec23p and phospho-Sec24p could only be detected *in vivo* when Hrr25p was expressed (Fig. 2c, lane 1).

As Sec23p binds to both TRAPPI and Hrr25p, we determined whether TRAPPI and Hrr25p compete for binding to Sec23p. To do this, the six subunits of the TRAPPI complex were co-expressed in bacteria and purified from bacterial lysates as described previously⁶. When similar amounts of purified His₆–Hrr25p and TRAPPI were incubated together, Hrr25p effectively competed with TRAPPI for binding to purified GST–Sec23p (Supplementary Fig. 4a, compare lanes 3–5 to GST controls in lanes 1 and 2). Binding was not dependent on kinase activity, as His₆–Hrr25p(K38A) bound as efficiently as His₆–Hrr25p (not shown). Because His₆–Hrr25p binds with higher affinity to GST–Sec23p ($K_d = 0.043 \pm 0.009 \mu$ M; Supplementary Fig. 4b) than TRAPPI ($K_d = 0.63 \pm 0.15 \mu$ M; Supplementary Fig. 4c), we determined whether it displaces TRAPPI that is pre-bound to GST–Sec23p. Increasing amounts of His₆–Hrr25p were mixed with GST–Sec23p beads pre-incubated with saturating amounts of TRAPPI (Fig. 2d). When the concentration of Hrr25p was increased (Fig. 2d, lanes 2–5), TRAPPI was released from the beads (Fig. 2d, top, lanes 2–5) into the supernatant (Fig. 2d, bottom, lanes 2–5) as His₆–Hrr25p bound to GST–Sec23p (Fig. 2d, middle). These findings show that Hrr25p and TRAPPI compete for binding to Sec23p, and indicate that Hrr25p could displace TRAPPI from Sec23p when COPII vesicles tether to the Golgi. Consistent with the possibility that phosphorylation of the coat blocks the binding of TRAPPI to Sec23p, we observed a decrease in the binding of TRAPPI to GST–Sec23p that contained phosphomimetic mutations at two phosphorylation sites (see later and Supplementary Fig. 5b).

Conservation of phosphorylation sites in Sec23p

Because Hrr25p phosphorylates Sec23p more efficiently than Sec24p, we focused on Sec23p for subsequent studies. Three Hrr25p phosphorylation sites were identified in Sec23p by mass spectrometry: T555, S742 and T747. Two of these phosphorylated residues, S742 and T747, are conserved from yeast to man and were analysed further (Fig. 3a, top). The T747 residue is a known Sar1p contact site, whereas S742 is within a disordered loop in the established structure of Sec23p

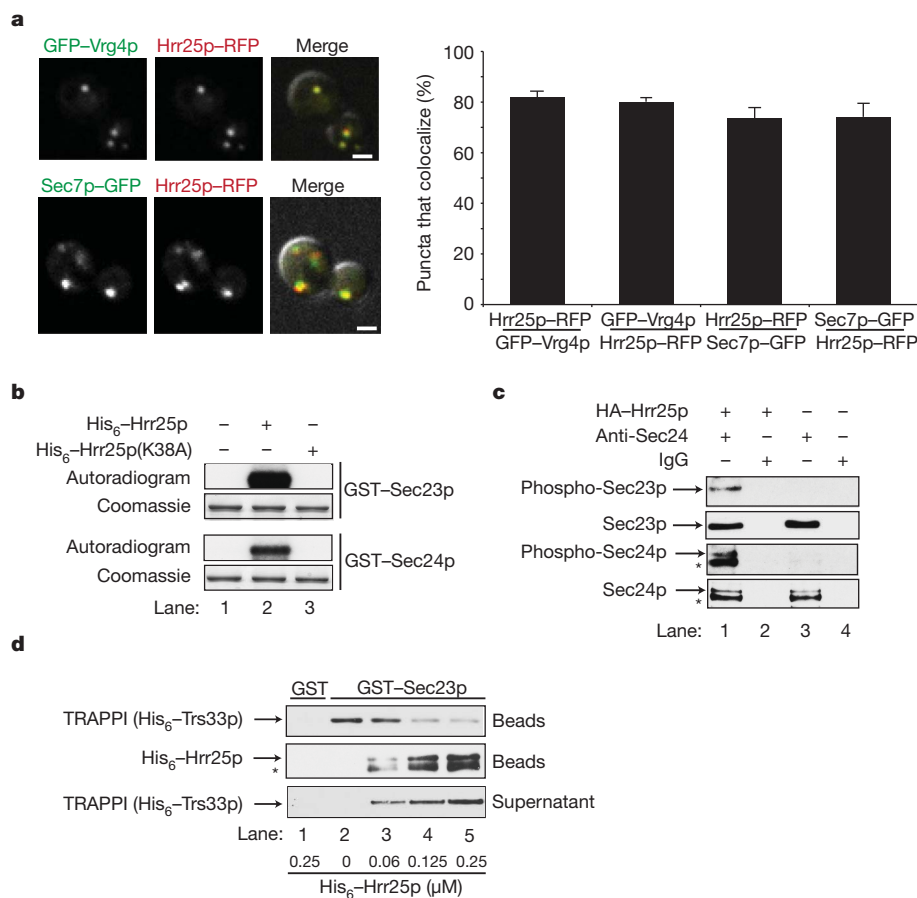


Figure 2 | Hrr25p resides on the Golgi and phosphorylates Sec23p/Sec24p. **a**, Hrr25p-RFP colocalizes with GFP-Vrg4p (top) and Sec7p-GFP (bottom). The green and red channels are merged with the differential interference contrast (DIC) image (right panel). The puncta that colocalize (s.d., $N = 3$ experiments) are shown on the right. Scale bar, 2 μm. **b**, GST-Sec23p and GST-Sec24p were incubated without (lane 1), or with His₆-Hrr25p (lane 2) or His₆-Hrr25p(K38A) (lane 3) and γ P³²-ATP. The autoradiogram and coomassie-stained gel are shown. **c**, Lysates expressing (lanes 1 and 2) or not expressing HA-Hrr25p (lanes 3 and 4) were immunoprecipitated with

anti-Sec24 antibody (lanes 1 and 3) or IgG (lanes 2 and 4) and immunoblotted with anti-phospho-Ser/Thr, anti-Sec23p and anti-Sec24p antibodies. **d**, TRAPPI, pre-bound to GST-Sec23p-containing beads (top), was incubated with increasing concentrations of His₆-Hrr25p. The beads were pelleted and the amount of TRAPPI in the supernatant (bottom) and pellet (top) was assessed. The Hrr25p that bound to the beads (middle) was also measured. TRAPPI, or Hrr25p, did not bind to GST (lane 1). The starred bands in **c** and **d** are degradation products of Sec24p and His₆-Hrr25p.

complexed with Δ23-Sar1p-GTP (PDB accession 2QTV)¹⁵. Initially, we used computational modelling to predict the consequences of phosphorylation at these sites. We added phosphates on S742 and T747 of Sec23p using Molsoft-ICM software and modified their orientations by several rounds of Monte Carlo optimization. Phosphorylation at T747 presented significant steric clashes with the surface of Sar1p in all possible orientations and was deemed incompatible with Sec23p binding to Sar1p-GTP (Fig. 3a, bottom). Phosphorylated S742 is also located at the Sec23p-Sar1p interface, but its location within a flexible loop limited our ability to predict consequences (Fig. 3a, bottom)¹⁵. We found that His₆-Δ23-Sar1p-GTPγS bound preferentially to GST-Sec23p (Supplementary Fig. 5a), and failed to bind GST-Sec23p harbouring the phosphomimetic S742D, T747E or S742D/T747E (ST-DE) mutations (Fig. 3b, top, lanes 1–4). Binding of His₆-Sec24p to GST-Sec23p was unaffected by the phosphomimetic mutations (Fig. 3b, bottom).

If Hrr25p phosphorylates Sec23p at Sar1p contact sites, the addition of His₆-Hrr25p to the transport assay should disrupt the binding of Sec23p to Sar1p-GTP and inhibit vesicle budding *in vitro*. When His₆-Hrr25p was added to the assay at the beginning of the reaction, budding was inhibited as the concentration of His₆-Hrr25p was increased (Fig. 3c). Inhibition was dependent on kinase activity, as no effect was seen with catalytically inactive His₆-Hrr25p(K38A)

(Fig. 3c). Consistent with this observation, a yeast strain harbouring the S742D/T747E mutations disrupted vesicle budding and fusion *in vitro* (Fig. 3d) and displayed a severe growth defect at 33 °C (Supplementary Fig. 5c). The defect in fusion may be the consequence of blocking the cycling of Sec23p on and off membranes (see Fig. 5a).

Because Hrr25p phosphorylates Sec23p at Sar1p contact sites, and Hrr25p competes with TRAPPI for binding to Sec23p, we wanted to address whether TRAPPI also competes with His₆-Δ23-Sar1p-GTPγS for binding to Sec23p. When we incubated a constant amount of His₆-Δ23-Sar1p-GTPγS with increasing amounts of TRAPPI, TRAPPI effectively competed with Δ23-Sar1p-GTPγS for binding to Sec23p (Fig. 4a, top). Similar results were obtained when increasing amounts of His₆-Δ23-Sar1p-GTPγS were incubated with a constant amount of TRAPPI (Fig. 4a, bottom). These results imply that TRAPPI and Sar1p-GTP bind to the same or overlapping site(s) on Sec23p. Consistent with this hypothesis, we found a decrease in the binding of TRAPPI to GST-Sec23p harbouring the S742D/T747E mutations (Supplementary Fig. 5b). Together, these findings imply that TRAPPI binds to Sec23p after Sar1p-GTP is released from membranes. This event seems to be conserved in higher eukaryotes (Fig. 4b), as we could only detect mBet3 on immuno-isolated tagged mammalian COPII vesicles (VSV-G-Myc) formed *in vitro* with GTP (–Sar1), but not the non-hydrolysable GTP analogue GMP-PNP (+Sar1) or control vesicles that lacked VSV-G-Myc.

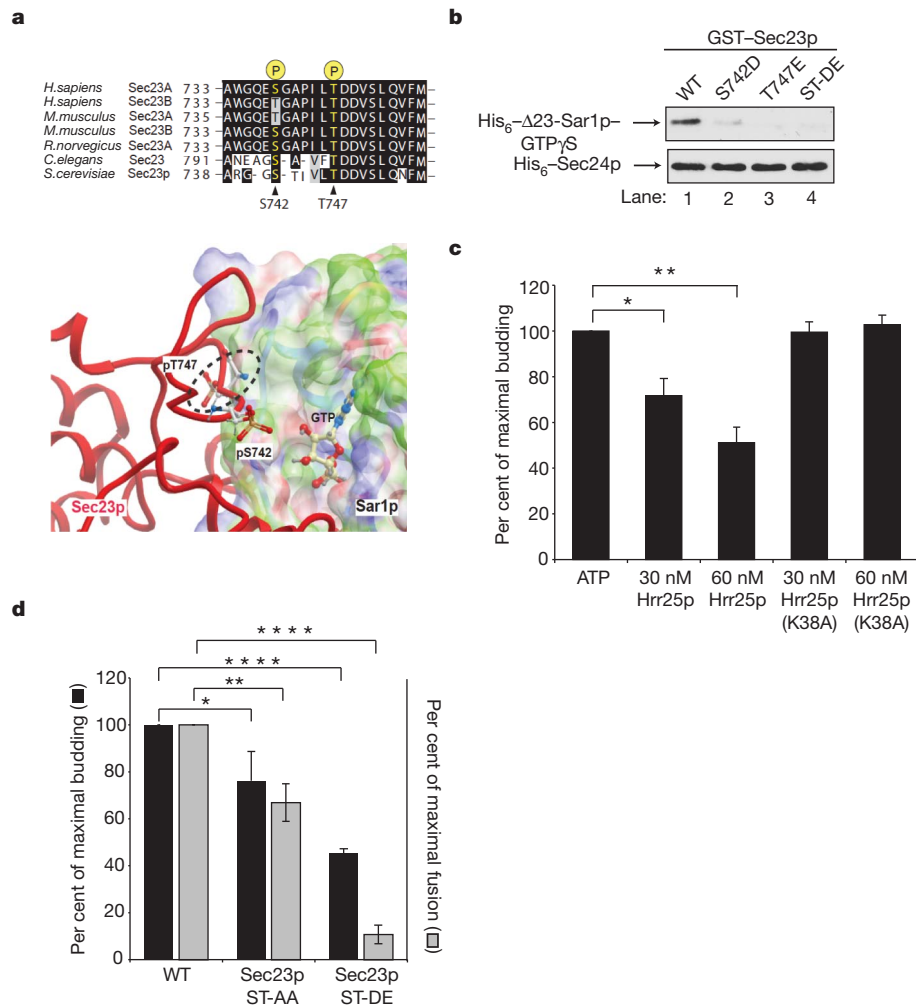


Figure 3 | Phosphorylation of S742 and T747 blocks ER-Golgi traffic *in vitro*. **a**, Top, the sequence flanking S742 and T747 in Sec23p was aligned with Sec23 orthologues. Bottom, phosphorylated S742 and T747 in Sec23p (red) are located at its interface with Sar1p. The electrostatic potential of the Sec23p-binding surface of Sar1p is coloured according to solvation properties of the residues (white, hydrophobic; green, polar; blue, basic; red, acidic). The dotted ellipsoid marks the steric clash between Sec23p and Sar1p-GTP. **b**, Top, wild-type (WT) and mutant GST-Sec23p fusion proteins were incubated with 10 nM of His₆-Δ23-Sar1p-GTPγS or His₆-Sec24p. A truncated form of Sar1p

was used for these studies because the full-length protein aggregates¹⁵. **c**, Increasing amounts of purified His₆-Hrr25p or His₆-Hrr25p(K38A) were added *in vitro* at the beginning of a complete transport reaction and vesicle budding was measured. Error bars represent s.d., *N* = 4. **d**, Fractions prepared from wild type and strains containing non-phosphorylatable (ST-AA) and phosphomimetic (ST-DE) Sec23p were assayed for vesicle budding and fusion as before⁶. Error bars represent standard error of the mean (s.e.m.), *N* = 4. **P* < 0.05, ***P* < 0.01, *****P* < 0.0001 Student's *t*-test.

Loss of Hrr25p activity inhibits fusion

To address the role of Hrr25p phosphorylation in ER-Golgi traffic *in vitro*, we used the ATP competitive inhibitor IC261, which selectively inhibits the highly conserved kinase domain of CK1δ^{11,16}. As seen in Fig. 4c, increasing concentrations of IC261 inhibited fusion. To address whether IC261 inhibits membrane fusion or vesicle tethering, we formed vesicles in the presence of Golgi with or without inhibitor, and then fractionated the reaction product on a sucrose velocity gradient that separates free vesicles (Supplementary Fig. 2b, top) from vesicles that are bound to the Golgi (Supplementary Fig. 2b, bottom). Subsequently, each fraction was treated with ConA Sepharose to precipitate radiolabelled pro-α-factor. Most of the vesicles bound to the Golgi in the presence of IC261, indicating that the inhibitor largely blocks fusion and not tethering (Supplementary Fig. 2b, bottom).

Interestingly, when the transport assay was performed with concentrations of IC261 that inhibited fusion, a stimulation in vesicle budding was observed (Fig. 4d). This finding indicates that dephosphorylation is needed for budding and is consistent with an earlier report showing that phosphorylated mammalian Sec24 cannot form a pre-budding complex⁹. A prediction of this result is that loss of kinase

activity should stimulate cargo export *in vivo* (see later) and could explain why a kinase loss-of-function mutation was identified as a suppressor of the *sec12-4* mutant¹². We were unable to address the role of CK1δ *in vitro* because the mammalian COPII vesicle tethering assay did not work robustly at the lower ATP concentrations needed to test the inhibitor.

To address whether phosphorylation and dephosphorylation alter the distribution of Sec23p on membranes *in vivo*, a differential fractionation experiment was performed with the conditional *hrr25* mutant after growth in galactose- or glucose-containing medium. The SNARE Bos1p served as a membrane marker for these studies. This analysis revealed the presence of a soluble pool of Sec23p when HA-Hrr25p was expressed (Fig. 5a, lanes 1–3). In the absence of HA-Hrr25p, however, all of the Sec23p was membrane-bound (Fig. 5a, lanes 4–6). Although Hrr25p activity seems to have a role in releasing Sec23p from membranes *in vivo*, we found it was not sufficient to uncoat the vesicles in the absence of Golgi membranes *in vitro* (data not shown).

In mammalian cells, COPII vesicles fuse to each other or with COPI (Golgi) vesicles to form a pre-Golgi compartment that matures into a

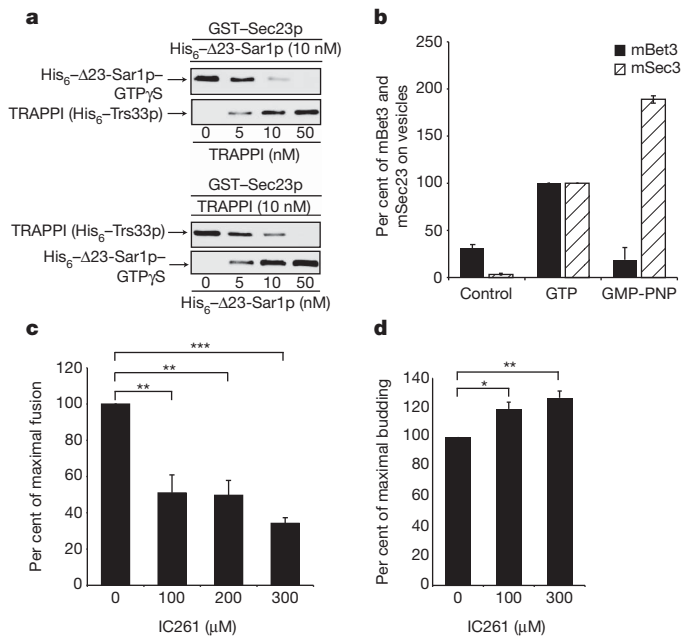


Figure 4 | TRAPPI and Sar1p-GTP cannot bind to Sec23p simultaneously. **a**, GST-Sec23p beads were incubated with 10 nM of His₆-Δ23Sar1p-GTPγS and increasing concentrations of TRAPPI (top), or 10 nM TRAPPI and increasing concentrations of His₆-Δ23Sar1p-GTPγS (bottom). **b**, mBet3 is on COPII vesicles formed with GTP (–Sar1), but not GMP-PNP (+Sar1). Vesicles formed *in vitro* with GTP or GMP-PNP were immunoprecipitated with an antibody against the cargo VSV-G-Myc and blotted for mBet3 and mSec23. The data was normalized to cargo yield (see Methods). *N* = 2, bars show the range. **c**, *In vitro* transport was performed with increasing concentrations of IC261 and fusion was measured. Error bars represent s.d., *N* = 4. **d**, The *in vitro* transport assay was performed as in Fig. 4c with IC261 and budding was measured. Error bars represent s.d., *N* = 3. **P* < 0.05, ***P* < 0.01, ****P* < 0.001 Student's *t*-test.

Golgi¹⁷. To address the role of CK1δ *in vivo*, we accumulated tsO45 VSV-G-GFP in the ER of NRK cells at 40 °C and then shifted the cells in the presence and absence of IC261 to 15 °C, a temperature that selectively slows traffic at the pre-Golgi compartment step¹⁸. In the presence of inhibitor, the pre-Golgi (marked by rbet1) but not early Golgi (marked by gpp130) was markedly dispersed (Fig. 5b). Consistent with the observation that inhibiting CK1δ function stimulates COPII vesicle budding and blocks fusion, VSV-G-GFP was more rapidly depleted from the ER and concentrated at peripheral sites of the pre-Golgi compartment (Fig. 5b, c), the site where COPII vesicles fuse. The VSV-G-GFP remained at the peripheral sites in the IC261-treated cells and failed to concentrate in the peri-Golgi region (Fig. 5d, e). Together, these findings imply that the events we describe here are conserved in higher cells.

Discussion

The CKI family of kinases represents a unique group of highly conserved serine/threonine kinases that regulate a variety of cellular processes, including membrane traffic^{11,12}. Here we report that Sec23p, a component of the inner shell of the COPII coat, sequentially interacts with three different binding partners, Sar1p, TRAPPI and Hrr25p, to control the direction of ER–Golgi traffic. These interactions define three different stages in vesicle traffic: budding, tethering and a pre-fusion step.

As Sar1p is required for fission^{19,20}, our findings imply that TRAPPI can only bind to Sec23p after vesicle fission and the release of Sar1p from membranes (Fig. 6, (1)). This ensures that COPII vesicle tethering is only initiated after a vesicle buds from the ER. Subsequently, TRAPPI activates Ypt1p on the vesicle (Fig. 6, (2)). Genetic studies and a kinetic analysis of GEF activity have revealed that TRAPPI is

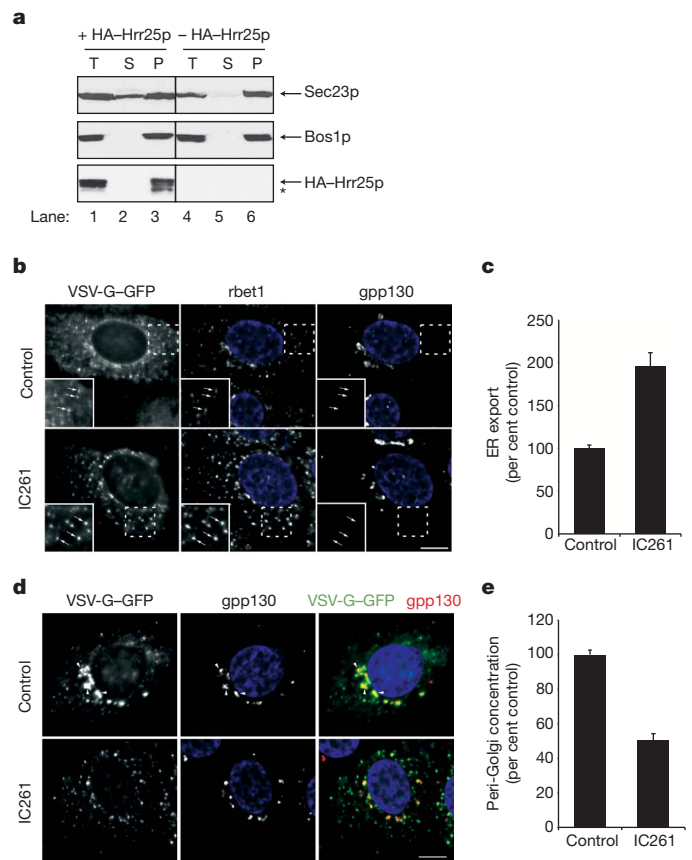


Figure 5 | In the presence of IC261, cargo is exported more rapidly but remains at peripheral pre-Golgi sites. **a**, The *hrr25* mutant was grown in galactose (lanes 1–3), or glucose (lanes 4–6) supplemented media for 10 h. Total (T) lysates were centrifuged at 150,000g and the supernatant (S) and pellet (P) fractions were analysed. The starred band is a degradation product of HA-Hrr25p. **b**, NRK cells that accumulated VSV-G-GFP in the ER at 40 °C were shifted to 15 °C for 30 min in the presence of DMSO (top) or 100 μM IC261 (bottom) to allow cargo transit from the ER to the ER–Golgi interface. Left, VSV-G-GFP fluorescence; middle, rbet1; right, gpp130. Insert in left corner is an expansion of the dotted box. Arrows, peripheral cargo-containing punctate structures that colocalize with rbet1 but not gpp130. **c**, Quantification, from Fig. 5b, of VSV-G-GFP fluorescence in pre-Golgi structures divided by VSV-G-GFP remaining in the ER (see Methods for calculation). Error bars represent s.e.m., *N* = 20 cells per bar, *P* < 0.0001 Student's *t*-test. **d**, Same as **b** only cells were incubated at 15 °C for 60 min. Left, VSV-G-GFP fluorescence; middle, gpp130; right, merge of VSV-G-GFP (green) and gpp130 (red). Arrowheads, cargo that concentrated in the Golgi area. **e**, Quantification, from Fig. 5d, of VSV-G-GFP fluorescence in punctate structures overlapping the Golgi divided by total fluorescence in punctate structures (see Methods for calculation). Error bars represent s.e.m., *N* = 20 cells per bar, *P* < 0.0001 Student's *t*-test. The nuclei are stained with DAPI. Scale bars, 10 μm.

more than a GEF²¹. GEFs typically release Rabs soon after they activate them. TRAPPI, however, forms a relatively stable ternary complex (TRAPPI–Ypt1p–nucleotide) with the Rab²¹, implying that it is also a Ypt1p effector. In parallel, the pool of Ypt1p–GTP that is released from TRAPPI can then recruit the long coiled-coil tether Uso1p (Fig. 6, (3)). When Uso1p bends, the vesicle comes into proximity with the Golgi, triggering the release of TRAPPI from the vesicle (Fig. 6, (3)). Hrr25p, which concentrates on the Golgi in yeast, could facilitate this release. Phosphorylation of the Sec23p/Sec24p complex by Hrr25p may be required, but is not sufficient for COPII vesicle uncoating. Another kinase could also be involved in this event as Sec31p, a known phosphoprotein²², seems to be phosphorylated by a different kinase¹² and Hrr25p cannot uncoat COPII vesicles *in vitro*.

Fusogenic SNARE motifs that bind to the coat must be unmasked before *trans*-SNARE complex formation can proceed at the target

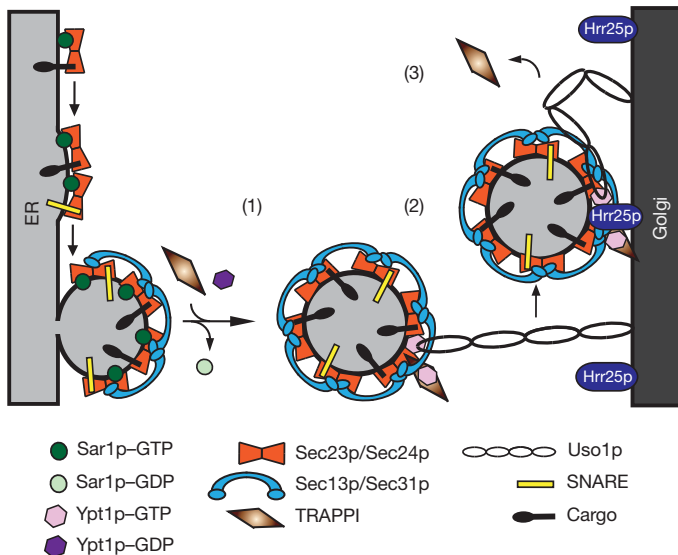


Figure 6 | Sec23p ensures the direction of ER–Golgi traffic. (1) TRAPPI binds to Sec23p after GTP is hydrolysed on Sar1p. (2) Ypt1p is activated by TRAPPI and Uso1p is recruited to the vesicle. Subsequently, Uso1p binds to the Golgi. (3) When Uso1p bends, it brings the vesicle to the Golgi. TRAPPI is then released from the vesicle and Hrr25p phosphorylates the Sec23p/Sec24p complex.

membrane³. Phosphorylation of Sec24p by Hrr25p/CKI δ could have a role in disengaging the SNAREs from the coat at the Golgi. As the COPII coat only acts in anterograde traffic⁴, the compartmentalization of a kinase that regulates membrane fusion at the Golgi, ensures an ER–Golgi v-SNARE will only pair with its cognate t-SNARE. The directionality imposed by this cycle also prevents the back-fusion of a COPII vesicle with the ER. The findings we report here describe a new role for Hrr25p/CKI δ that may extend to other CKI family members and coats.

METHODS SUMMARY

Yeast COPII vesicles were formed *in vitro* with donor cells, cytosol, with or without Golgi for 90 min at 20 °C or 27 °C or 120 min at 17 °C as described previously⁶. The vesicle binding assays were performed as described previously⁶. Additional information is provided in Methods.

Protein and antibody purifications were performed as before⁶. Mass spectrometry analysis, *in vitro* bindings assays, kinase and immunoprecipitation assays, microscopy, the construction of phosphomimetic mutations and all studies with mammalian cells are described in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 September 2010; accepted 23 February 2011.

Published online 1 May 2011.

1. Jahn, R. & Scheller, R. H. SNAREs—engines for membrane fusion. *Nature Rev. Mol. Cell Biol.* **7**, 631–643 (2006).
2. Lewis, M. J., Rayner, J. C. & Pelham, H. R. A novel SNARE complex implicated in vesicle fusion with the endoplasmic reticulum. *EMBO J.* **16**, 3017–3024 (1997).

3. Mossessova, E., Bickford, L. C. & Goldberg, J. SNARE selectivity of the COPII coat. *Cell* **114**, 483–495 (2003).
4. Lee, M. C. S., Miller, E. A., Goldberg, J., Orci, L. & Schekman, R. Bi-directional protein transport between the ER and Golgi. *Annu. Rev. Cell Dev. Biol.* **20**, 87–123 (2004).
5. Whyte, J. R. C. & Munro, S. Vesicle tethering complexes in membrane traffic. *J. Cell Sci.* **115**, 2627–2637 (2002).
6. Cai, H. *et al.* TRAPPI tethers COPII vesicles by binding the coat subunit Sec23. *Nature* **445**, 941–944 (2007).
7. Barlowe, C. *et al.* COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. *Cell* **77**, 895–907 (1994).
8. Flanagan, J. J. & Barlowe, C. Cysteine-disulfide cross-linking to monitor SNARE complex assembly during endoplasmic reticulum–Golgi transport. *J. Biol. Chem.* **281**, 2281–2288 (2006).
9. Dudognon, P., Maeder-Garavaglia, C., Carpentier, J. & Paccaud, J. Regulation of a COPII component by cytosolic O-glycosylation during mitosis. *FEBS Lett.* **561**, 44–50 (2004).
10. Milne, D. M., Looby, P. & Meek, D. W. Catalytic activity of protein kinase CK1 δ (casein kinase 1 δ) is essential for its normal subcellular localization. *Exp. Cell Res.* **263**, 43–54 (2001).
11. Yu, S. & Roth, M. G. Casein kinase I regulates membrane binding by ARF GAP1. *Mol. Biol. Cell* **13**, 2559–2570 (2002).
12. Murakami, A., Kimura, K. & Nakano, A. The inactive form of a yeast casein kinase I suppresses the secretory defect of the sec12 mutant. *J. Biol. Chem.* **274**, 3804–3810 (1999).
13. Lusk, C. P. *et al.* Nup53p is a target of two mitotic kinases, Cdk1p and Hrr25p. *Traffic* **8**, 647–660 (2007).
14. Kafadar, K. A., Zhu, H., Snyder, M. & Cyert, M. S. Negative regulation of calcineurin signaling by Hrr25p, a yeast homolog of casein kinase I. *Genes Dev.* **17**, 2698–2708 (2003).
15. Bi, X., Corpina, R. A. & Goldberg, J. Structure of the Sec23/24–Sar1 pre-budding complex of the COPII vesicle coat. *Nature* **419**, 271–277 (2002).
16. Mashhoon, N. *et al.* Crystal structure of a conformation-selective casein kinase-1 inhibitor. *J. Biol. Chem.* **275**, 20052–20060 (2000).
17. Xu, D. & Hay, J. C. Reconstitution of COPII vesicle fusion to generate a pre-Golgi intermediate compartment. *J. Cell Biol.* **167**, 997–1003 (2004).
18. Schweizer, A. *et al.* Identification of an intermediate compartment involved in protein transport from endoplasmic reticulum to Golgi apparatus. *Eur. J. Cell Biol.* **53**, 185–196 (1990).
19. Bielli, A. *et al.* Regulation of Sar1 NH₂ terminus by GTP binding and hydrolysis promotes membrane deformation to control COPII vesicle fission. *J. Cell Biol.* **171**, 919–924 (2005).
20. Lee, M. C. S. *et al.* Sar1p N-terminal helix initiates membrane curvature and completes the fission of a COPII vesicle. *Cell* **122**, 605–617 (2005).
21. Chin, H. F. *et al.* Kinetic analysis of the guanine nucleotide exchange activity of TRAPP, a multimeric Ypt1p exchange factor. *J. Mol. Biol.* **389**, 275–288 (2009).
22. Salama, N. R., Chuang, J. S. & Schekman, R. W. Sec31 encodes an essential component of the COPII coat required for transport vesicle budding from the endoplasmic reticulum. *Mol. Biol. Cell* **8**, 205–217 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Schekman, M. Cyert, L. Miller, B. Glick, M. Lowe and E. Mizuno-Yamasaki for strains, constructs, purified protein and antibody; K. Reinisch for discussions; S. Chen, H. Cai and M. Garcia-Marcos for advice; W. Zhou and A. Loughheed for technical assistance. This work was supported by the Howard Hughes Medical Institute. Salary support for S.F.-N., D.B. and S.M. was provided by the Howard Hughes Medical Institute, P.G. was funded by the Burroughs Wellcome Fund and M.G. by the SRP Super Fund Research Program. Work at the University of Montana was supported by National Institutes of Health grant GM-059378 (to J.C.H.) and the National Institutes of Health COBRE Center grant RR-015583.

Author Contributions C.L., D.B., S.M., D.N. and J.H. performed experiments and analysed data. M.G. did the mass analysis and P.G. did the computational modelling and analysed data. S.F.-N. directed the project, analysed data and wrote the paper. C.L., D.B., J.H. and P.G. co-wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.F.-N. (sfnovick@ucsd.edu).

METHODS

Yeast *in vitro* transport and vesicle binding assays. For the vesicle binding assays, the permeabilized cells were pelleted after the *in vitro* transport reaction as described previously⁶. The conditions used to remove the cells did not pellet slowly sedimenting membranes that contain radiolabelled cargo. The supernatant was transferred to a new tube that contained 40 µl of a 50% slurry of TRAPPI-containing beads⁶. The final volume of the reaction was adjusted to 500 µl with TBPS (25 mM HEPES (pH 7.2), 115 mM potassium acetate, 2.5 mM magnesium acetate, 250 mM sorbitol plus protease inhibitors) and the reaction was incubated for 2 h at 4 °C. The beads were washed three times with 750 µl of TBPS and counted. The counts from a no vesicle control reaction (reaction with apyrase with or without Golgi) were subtracted as background. The binding of vesicles formed with cytosol was considered to be 100%, and the amount of binding (with or without Golgi) was adjusted to equal Concanavalin A (ConA) counts (equal vesicles). Two-stage transport assays were also performed. Vesicles were formed in the absence of Golgi, the cells were pelleted and an equal number of vesicles were incubated with or without Golgi before the binding assay was performed.

For reactions with IC261, the ATP concentration of the 10× ATP stock was lowered to 1.7 mM ATP. For the vesicle tethering assay, free vesicles were separated from vesicles that bound to the Golgi on a sucrose velocity gradient as described previously⁶.

***In vitro* kinase assay.** Purified GST, GST–Sec23p and GST–Sec24p (5 µg), immobilized on beads, were incubated with 250 ng of His₆–Hrr25p or catalytically inactive His₆–Hrr25p(K38A) or no kinase in kinase assay buffer (50 mM HEPES pH 7.4, 2 mM EDTA, 10 mM MgCl₂, 1 mM DTT, 5 mM cold ATP, 2.5 µCi γP³²-ATP, 100 µM sodium orthovanadate, 10 mM sodium fluoride, 10 mM sodium pyrophosphate and protease inhibitors) for 1 h at 30 °C. The beads were washed twice with 1× PBS and eluted in 25 µl of sample buffer by heating to 100 °C for 5 min. The samples were analysed by autoradiography.

Nucleotide loading of His₆–Δ23–Sar1p. His₆–Δ23–Sar1p was loaded with the desired nucleotide (GDP or GTPγS) overnight at 4 °C in the following buffer: 20 mM HEPES pH 7.2, 150 mM NaCl, 1 mM MgCl₂ with 0.1 mM of nucleotide.

***In vitro* binding assays with recombinant proteins.** For the binding experiments with the phosphomimetic mutations in GST–Sec23p, equimolar amounts (0.1 µM) of GST, GST–Sec23p beads with or without phosphomimetic mutations were incubated with either 5–10 nM of His₆–Δ23–Sar1p–GTPγS (nucleotide was loaded as described earlier) or 10 nM of purified TRAPPI⁶ in binding buffer I (25 mM HEPES pH 7.2, 150 mM NaCl, 2% Triton X-100, 1 mM DTT, 2 mM EDTA, 0.5 mM MgCl₂ and protease inhibitors) for 3–4 h at 4 °C. The beads were washed three times with binding buffer I and eluted in 25 µl of sample buffer by heating to 100 °C for 5 min.

For the TRAPPI displacement assay, 0.5 µM of TRAPPI was incubated overnight at 4 °C in binding buffer I with 0.1 µM of GST–Sec23p or GST. The next day, the beads were washed three times in binding buffer I and resuspended in binding buffer II (25 mM HEPES pH 7.2, 150 mM NaCl, 0.1% Triton X-100, 1 mM DTT, 2 mM EDTA, 0.5 mM MgCl₂ and protease inhibitors). Increasing amounts (0–0.25 µM) of purified His₆–Hrr25p were added to the reactions and incubated for 4 h at 4 °C. Subsequently, the supernatants were aliquoted into fresh tubes and heated in sample buffer to 100 °C for 5 min. The beads were washed three times with binding buffer I and eluted as described earlier.

For the competition assay between Hrr25p and TRAPPI, 0.3 µM of GST–Sec23p or GST beads were incubated with 0.25 µM of purified TRAPPI and/or 0.15 µM of His₆–Hrr25p in binding buffer I for 3–4 h at 4 °C. For the competition assay between Sar1p and TRAPPI, 0.1 µM of GST–Sec23p or GST beads were incubated with 10 nM of purified His₆–Δ23Sar1p–GTPγS with increasing concentrations (0–50 nM) of TRAPPI as above. For the reciprocal experiment, GST fusion proteins were incubated with 10 nM of TRAPPI and increasing concentrations (0–50 nM) of His₆–Δ23Sar1p–GTPγS. The beads were then washed three times with binding buffer I and eluted as above. The molarity of TRAPPI was calculated based on the amount of Trs33p in the complex.

Yeast immunofluorescence microscopy. Cells expressing Vrg4p–GFP or Sec7p–GFP and Hrr25p–RFP were grown to an OD_{599 nm} of 0.5–1.5 in YPD medium. One to two OD_{599 nm} units were pelleted and resuspended in 25 µl of YPD medium. Cells were examined with a Carl Zeiss Observer Z.1 spinning-disk confocal fluorescence microscope using DIC, GFP, or RFP filters with a ×100 oil-immersion objective. Images were captured with a Zeiss AxioCam MRm and analysed using AxioVision Rel. 4.7 software. At least 300 puncta (and 100 cells) were examined in three separate experiments that were used to calculate the s.d. shown in Fig. 2a.

Growth conditions for the *hrr25* mutant. SFNY1941 (*MATa ura3-52 lys2-801 ade2-101 trpΔ63 his3-Δ200 leu2-Δ1 hrr25Δ1::loxP-kanMX-loxP pKK204(2µ pGAL-3HA-HRR25^{degron})*) was grown in YP–Raf–Gal (2% Raffinose, 0.5% Galactose) medium to OD_{599 nm} = 1–2. A total of 500 OD_{599 nm} units of cells were pelleted

under sterile conditions and shifted to either YP–Raf–Gal or YPD (YP + 2% glucose) medium for 10 h.

Differential centrifugation experiment. A total of 100 OD_{599 nm} units of cells were pelleted, resuspended in 2 ml of spheroplast buffer (1.4 M sorbitol, 100 mM sodium phosphate pH 7.5, 0.35% 2-mercaptoethanol and 0.5 mg ml^{−1} zymolyase) and incubated for 30 min at 37 °C. The spheroplasted cells were then divided into four 0.5 ml aliquots and centrifuged over a 1 ml sorbitol cushion (1.7 M sorbitol, 100 mM HEPES pH 7.2) for 5 min at 3,800g at 4 °C in a microfuge. The supernatant was removed and the four pellets were resuspended in 1 ml of lysis buffer (100 mM HEPES pH 7.2, 1 mM EGTA, 0.2 mM DTT, 1 mM PMSF and protease inhibitors) and lysed using a Dounce homogenizer. The lysate was centrifuged for 2 min at 500g at 4 °C in a microfuge and the supernatant was transferred to a new tube. An aliquot of this fraction (100 µl) was mixed with 50 µl of 3× sample buffer (total fraction, T) and heated at 100 °C for 5 min, while the remaining portion (600 µl) was centrifuged for 90 min at 190,000g at 4 °C in a Beckman SW55 Ti rotor. The lipid layer was removed and 100 µl of the supernatant (S) was mixed with 50 µl of 3× sample buffer and heated at 100 °C for 5 min. The pellet (P) was resuspended in 500 µl of lysis buffer and 100 µl was mixed with 50 µl of 3× sample buffer and heated at 100 °C for 5 min.

Immunoprecipitation assay to detect phosphorylation. SFNY1941 was grown as described earlier and whole-cell lysates (10 mg) were immunoprecipitated with anti-Sec24 antibody. The immunoprecipitates were then immunoblotted with anti-phospho-Ser/Thr (BD Biosciences, 1:500 dilution), anti-Sec24 (1:1,000 dilution) and anti-Sec23 antibodies (1:1,000 dilution).

Mass spectrometry analysis. *In vitro* phosphorylated GST–Sec23p was trypsin-digested and subjected to liquid chromatography coupled with tandem mass spectrometry (LC–MS/MS) analysis as described previously²³.

Generation of Sec23p phosphomimetic mutations. The phosphomimetic mutations (S742D, T747E and S742D/T747E) in GST–Sec23p were generated by the two-step PCR method for site-directed mutagenesis using pPE124 (ref. 24) as the template. Mutations in pRS414-SEC23(S742D/T747E) were generated on pCF364 (ref. 25) using the QuikChange Site-directed mutagenesis kit (Agilent Technologies). All constructs were verified by sequencing.

Analysis of mBet3 and mSec23 on mammalian COPII vesicles. A detailed description of the generation, immunoisolation and immunoblotting of COPII transport intermediates derived from semi-intact normal rat kidney (NRK) cells is described elsewhere¹⁷. Briefly, a VSV–G–Myc construct was introduced by electroporation and its expression in the ER was amplified using vaccinia virus VTF-7 at 41 °C. After permeabilization, the VSV–G–Myc-expressing cells or control untransfected NRK cells were suspended in a vesicle budding cocktail and incubated at 32 °C for 30 min. Subsequently, the donor cells were removed by sedimentation. For the p115 experiment, the supernatant (which contains released COPII vesicles) was then incubated with purified full-length His₆–p115 (ref. 26) for 60 min at 32 °C. The p115 preparation was functional in tests of interactions with ER–Golgi SNAREs (not shown). The suspension of transport intermediates was then subjected to immunoisolation using anti-Myc antibody. Proteins were eluted from the beads using 0.1 M glycine pH 2.5, neutralized, concentrated, and analysed on a 4–20% gradient SDS polyacrylamide gel followed by western blot analysis. To quantitate the abundance of mBet3 and mSec23 on the isolated vesicles and to normalize to vesicle yield, we divided the band intensity by the signal for the cargo marker VSV–G–Myc and syntaxin 5 for each lane. The cargo-normalized mBet3 and mSec23 signals were then expressed as a percentage relative to the GTP condition. GMP–PNP was used at 100 µM concentration.

Analysing ER–Golgi traffic *in vivo* in the presence of IC261. NRK cells were electroporated with a plasmid encoding VSV–G ts045–GFP, plated on glass coverslips in 6-well plates and incubated overnight at 40 °C. Ten minutes before the temperature shift, 100 µM IC261 (solubilized in DMSO) or DMSO was added to the medium at 40 °C. After 10 min, the coverslips were either fixed in 4% paraformaldehyde for 30 min, or shifted to 15 °C medium containing IC261 or DMSO. At 15 °C, VSV–G–GFP can leave the ER, but accumulates in swollen peripheral ER–Golgi interface structures that only slowly move towards the Golgi area²⁷. The 15 °C treatment makes ER exit, pre-Golgi assembly and transport to the Golgi more resolvable. Coverslips were fixed after 30 or 60 min at 15 °C.

After fixation, coverslips were treated twice for 10 min with 0.1 M glycine and then the samples were permeabilized in BSA/goat serum blocking solution containing 0.35% saponin. All subsequent antibody incubations and washes were carried out in blocking solution containing saponin. Primary antibody incubations included the anti-rbet1 mouse monoclonal antibody 16G6 (ref. 28) and a rabbit polyclonal antiserum against gpp130 (Covance Research Products). 16G6 is known to label rbet1 more intensely when the antigen accumulates at peripheral sites²⁹. Secondary antibodies were goat anti-mouse-cy3 and goat anti-rabbit-cy5. DAPI was also included during the secondary antibody incubation. After extensive

washing, the coverslips were mounted in Slow-fade Gold mounting medium (Invitrogen) and imaged using the wide-field microscope and instrumentation methods described before²⁹. Briefly, each field of cells were captured in four colours (GFP, cy3, cy5 and DAPI) at 21 focal planes through the sample. Image stacks were then deconvolved using the Huygens algorithm (Scientific Volume Imaging). Maximum intensity projections from five consecutive image planes were used for quantification and display.

To quantify the transfer from the ER to pre-Golgi structures, each set of images was opened as an image stack in the Openlab program (Improvision). The extracellular background was subtracted from each image, and then a roughly square region of interest (ROI) that abutted the nucleus and extended approximately three-quarters of the distance to the edge of the cell was hand-drawn. This ROI was chosen such that it constituted a region of cytoplasm completely free of any Golgi labelling. The rbt1 image was used to define a pre-Golgi mask within this larger ROI using an intensity threshold of 4–6× the rbt1 labelling background. The pre-Golgi mask was then subtracted from the original ROI to create an ER mask. The pre-Golgi and ER mask were sequentially applied to the VSV-G–GFP image to determine the average maximum intensity of cargo spots in the pre-Golgi compartment, and the average intensity of cargo in the ER, respectively. The ratio of these two parameters, average maximum intensity in the pre-Golgi compartment divided by average intensity in the ER, was determined for each of the twenty randomly sampled images from each condition.

To quantify the concentration of pre-Golgi structures near the Golgi, a Golgi mask was derived from the gpp130 image using an intensity threshold of 10% of maximum intensity for the Golgi area of interest. A total punctate cargo mask was derived from the VSV-G–GFP image by choosing an intensity threshold for each cell sample such that a faint punctate spot would be captured in the mask but residual diffuse ER labelling would not. This cargo mask was then superimposed on the Golgi mask and all cargo-containing objects that did not partially or

completely overlap with a Golgi object were deleted from the cargo mask. This peri-Golgi cargo mask, as well as the original total cargo mask were sequentially applied to the VSV-G–GFP image to determine the total intensity of cargo spots in the peri-Golgi region and in the whole cell, respectively. The ratio of these two parameters, total intensity of cargo spots in the peri-Golgi region divided by total intensity of cargo spots in the cell, was determined for each of the 20 randomly sampled images analysed for each condition.

For both quantifications, the average raw ratio for DMSO-treated cells from a given experiment was defined as 100% and each individual ratio value from that experiment was expressed relative to 100%. This normalization step allowed combination of quantifications from independent experiments to produce the values for Fig. 5c and e.

All experiments in the manuscript were performed at least three times or more on separate days.

23. Guttman, M. *et al.* Interactions of the NPXY microdomains of the low density lipoprotein receptor-related protein 1. *Proteomics* **9**, 5016–5028 (2009).
24. Gimeno, R. E., Espenshade, P. & Kaiser, C. A. COPII coat subunit interactions: Sec24p and Sec23p bind to adjacent regions of Sec16p. *Mol. Biol. Cell* **7**, 1815–1823 (1996).
25. Fromme, J. C. *et al.* The genetic basis of a craniofacial disease provides insight into COPII coat assembly. *Dev. Cell* **13**, 623–634 (2007).
26. Diao, A., Frost, L., Morohashi, Y. & Lowe, M. Coordination of golgin tethering and SNARE assembly: GM130 binds syntaxin 5 in a p115-regulated manner. *J. Biol. Chem.* **283**, 6957–6967 (2008).
27. Saraste, J. & Svensson, K. Distribution of the intermediate elements operating in ER to Golgi transport. *J. Cell Sci.* **100**, 415–430 (1991).
28. Hay, J. C. *et al.* Localization, dynamics, and protein interactions reveal distinct roles for ER and Golgi SNAREs. *J. Cell Biol.* **141**, 1489–1502 (1998).
29. Bentley, M. *et al.* Vesicular calcium regulates coat retention, fusogenicity, and size of pre-Golgi intermediates. *Mol. Biol. Cell* **21**, 1033–1046 (2010).